Fundamentals of Behavioral Statistics

Second Edition

Runyon-Haber

FUNDAMENTALS OF BEHAVIORAL STATISTICS

Second Edition

RICHARD P. RUNYON, C. W. Post College
AUDREY HABER, University of California, Los Angeles

ADDISON-WESLEY PUBLISHING COMPANY
Reading, Massachusetts · Menlo Park, California · London · Don Mills, Ontario

This book is in the
ADDISON-WESLEY SERIES IN PSYCHOLOGY

 Printed in the United States of America. Published simultaneously in Canada. Library of Congress Catalog Card No. 79-136126.

Preface to Second Edition

We should like to express our gratitude for the excellent reception of our first edition by instructors using the text in the classroom. Many of our colleagues have taken time from busy schedules to send comments and suggestions, many of which we have incorporated into this present edition.

In rewriting the text to include new material, we have made many improvements in the presentation of topics previously covered. We have both consolidated and expanded. We have attempted to present certain topics in a more concise and economical fashion. In other instances, however, we have elaborated on our discussion in an attempt to help clarify some of the more difficult concepts. We have added an optional section on set theory and its relationship to probability. The analysis of variance chapter has been rewritten to include a general discussion of multiple comparison tests and a specific example employing Tukey's HSD test.

In addition to expanding the exercises at the end of each chapter, we have prepared a student workbook which includes detailed program reviews, hundreds of true-false and multiple-choice items and additional work problems. While we have introduced a number of changes in this edition, we have attempted to retain many of the features that were so well-received in the previous edition.

A number of individuals have directly or indirectly contributed to the present edition. Specifically, we should like to express our appreciation to Lois Runyon for her cheerful acceptance of the many disruptions of her household during the preparation of this manuscript. In addition, we should like to extend our thanks to Jerry Jassenoff for his patience and understanding during this hectic period.

R.P.R.
A.H.

Greenvale, L.I., N.Y.
Los Angeles, Calif.
October 1970

Preface to First Edition

In our years of teaching introductory statistics to students in the behavioral sciences, we have been guilty of the crime—if it is a crime—of switching textbooks with such frequency that our appearance in the college book store has become a source of acute discomfort to the manager of the establishment (where profits are derived largely from the resale of used texts). Our experiences in this connection have not been unique; we have found that discussion with many of our colleagues in other institutions reveals similar difficulties in settling on a single text in statistics. These discussions have divulged two frequent sources of complaint.

1. Current texts tend to vary between those which are so oversimplified that they insult the intelligence of the student, and others which presume more mathematical sophistication than is found among most undergraduate students in the behavioral sciences.

2. Many of the current texts do not reflect the latest advances in small sample statistics or the more general availability of high-speed automatic calculators on college campuses.

To begin with, we are strongly opposed to texts which feature a "cookbook" approach to statistical analyses. Besides insulting the intelligence of the student, such texts tend to become the behavioral scientist's "Merck Manual" or a young student's "Dr. Spock." This type of text may easily lead the student into concluding that statistics is a Procrustean bed in which one demonstrates his capacity to withstand the pain of memorizing esoteric formulas and the various rules for their application. On the other hand, we recognize that a strictly mathematical approach to introductory statistics leaves many otherwise capable students floundering in a morass of mathematical symbols and abstractions which have little relevance to practical applications. We have attempted to effect a compromise between these two approaches. Thus although we have usually selected simple examples to illustrate the calculation of the various statistics, we have not hesitated to include algebraic proofs when such

proofs are within the grasp of most students of the behavioral sciences and, additionally, when they demonstrate some fundamental relationship.

On the second point, the failure of many current texts to reflect the latest statistical advances and technological change, we have made a deliberate effort to eliminate all materials which, from our experience, are rarely employed on the contemporary scene. Thus coded score methods, which were vital in the days of large sample statistics and before the advent of the high speed calculators, will not be found in the text. While some instructors, in a moment of acute nostalgia, may lament the loss of this relic from the "horse and buggy" days of statistics, we feel that instruction in the coded score methods is wasteful of time, complicates the life of the harried student who comes to believe that knowledge of formulas is the essence of statistics, and advances no fundamental insights into the nature of statistical analysis. The use of the correlational charts has suffered a similar death for the same reason. To critics, we may only point out that we have never, in our research, had occasion to employ coded score methods or the correlational chart, and we have never met anyone who has!

On the other hand, we have not hesitated to introduce new statistical techniques which we feel represent an advance over prior methods. Thus the Sandler A-statistic, which is algebraically equivalent to the Student t-ratio with correlated samples, has been introduced because it drastically reduces the computational procedures required to arrive at a statistical decision.

Furthermore, we have attempted to preserve the distinction between population parameters and sample statistics in the testing of hypotheses. Since our approach is consistent throughout the text, we hope to eliminate some of the confusion that sometimes arises in this context.

A word about the organization of the text. Many recent statistics textbooks have relegated descriptive statistics to a place of secondary importance. While it is not our contention that descriptive statistical techniques represent anything beyond the fundamentals of statistical analysis, a mastery of these techniques is prerequisite to the understanding and application of the concepts and procedures involved in inferential statistics. We have attempted, throughout the text, to demonstrate the continuity of these two branches of statistics.

Secondly, the statistical tables in Appendix III have been carefully prepared to minimize the student's difficulties in his use of them. For example, most tables are preceded by a brief description of the procedures involved in their application. In addition, wherever appropriate, critical values for rejecting the null hypothesis are shown in terms of one- and two-tailed values at various levels of significance. Finally, some tables have been reduced in complexity, making it possible to locate the relevant information in a shorter period of time with less chance of error. For example, the Mann-Whitney U-statistic is commonly shown in ten different tables. To further complicate matters, some of the tables provide exact probability values, whereas others provide critical values for rejecting the null hypothesis at various levels of significance. We have reduced these tables to four, showing only the critical values required to reject the null hypothesis. Furthermore, we have expressed these values

in terms of both U and U' so that the student need not be concerned about which of these statistics he has calculated when he employs the Mann-Whitney test.

The exercises at the end of each chapter are an extremely important and integral part of the text since, in addition to illustrating fundamental relationships, they require the student to formulate many significant statistical concepts himself.

Finally, the book has been organized so that the first fifteen chapters constitute, in our opinion, a thorough introduction to the fundamentals of descriptive and inferential statistics. For the instructor desiring more advanced statistical procedures, we have included, in the last four chapters of the text, such topics as analysis of variance, the power and power efficiency of a statistical test, and several of the more widely employed nonparametric tests of significance.

In brief, then, we have attempted to produce a textbook for a one-semester introductory course in statistics for students in the behavioral sciences. It is our hope that the student will gain an appreciation of the usefulness of the statistical method in his professional field, that he will have a good understanding of the assumptions and logic underlying the application of the statistical tools, that he will be able to select the appropriate statistical technique and perform the necessary computations, and, finally, that he will know how to interpret and understand the results of his efforts.

We are grateful to many people who have contributed to this book. We express our sincere appreciation to Dr. Nancy Wiggins for her many excellent and stimulating comments delivered at varying stages of the preparation of the manuscript. In addition, we are deeply indebted to Millicent Cowit, Ruth DeMarco, Norma Morrow, and Fleeta Runyon for their painstaking efforts in typing a manuscript which, because of the many unfamiliar symbols and formulas, must have produced many moments of anxiety and frustration.

Finally, we wish to express our gratitude to the many authors and publishers who have permitted us to adapt or reproduce material originally published by them. We have cited each source wherever it appears. We are indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Cambridge, to Dr. Frank Yates, F.R.S., Rothamsted and to Messrs. Oliver and Boyd, Ltd., Edinburgh, for permission to reprint and adapt tables from their books, *Statistical Tables for Biological, Agricultural, and Medical Research* and *Methods for Research Workers*.

Greenvale, L. I., N. Y.
January 1967

R.P.R.
A.H.

Contents

Section II INFERENTIAL STATISTICS

A. Parametric Tests of Significance

Section I
Descriptive Statistics

The Definition of Statistical Analysis

1

1.1 INTRODUCTION

If we were to ask the "man on the street" what statistics means to him, we would, in all likelihood, obtain some answers such as, "Statistics is 'hocus pocus' with numbers. By manipulating these numbers according to certain secret and well-guarded rules, we can prove anything we have a mind to." Or, "Statistics is the refuge of the uninformed. When we can't prove our point through the use of sound reasoning, we fall back upon statistical 'mumbo-jumbo' to confuse and demoralize our opponents." Or "Statistics is merely a collection of facts. Statisticians concern themselves with such vital issues as the number of bath tubs in the State of Kentucky in 1929, the number of men who grow mustaches to irritate their spouses and the number of wives who retaliate by growing beards."

It is true that all these activities, and more, are widely attributed to the field of statistics. It is *not* true, however, that statisticians engage in them. What, then, is statistics all about? Although it would be virtually impossible to obtain a general consensus on the definition of statistics, it is possible to make a distinction between two definitions of statistics.

1. Statistics is commonly regarded as a *collection* of numerical facts which are expressed in terms of summarizing statements and which have been collected either from several observations or from other numerical data. From this perspective, statistics constitutes a collection of statements such as, "The average I.Q. of 8th grade children is . . . ," or "Seven out of ten people prefer Brand X to Brand Y," or "The New York Yankees hit 25 home runs over a two week span during . . ."

2. Statistics may also be regarded as a *method* of dealing with data. This definition stresses the view that statistics is a tool concerned with the collection, organization, and analysis of numerical facts or observations.

The second definition constitutes the subject matter of this text.

A distinction may be made between the two functions of the statistical method: *descriptive* statistical techniques and *inferential* or *inductive* statistical techniques.

The major concern of descriptive statistics is to present information in a convenient, usable and understandable form. Inferential statistics, on the other hand, is concerned with generalizing this information, or, more specifically, with making inferences about populations which are based upon samples taken from the populations.

In describing the functions of statistics, certain terms have already appeared with which you may or may not be familiar. Before elaborating on the differences between descriptive and inductive statistics, it is important to learn the meaning of certain terms which will be employed repeatedly throughout the text.

1.2 DEFINITIONS OF COMMON TERMS USED IN STATISTICS

Variable: a characteristic or phenomenon which may take on different values. Thus, weight, I.Q., and sex are variables since they will take on different values when different individuals are observed. A variable is contrasted with a constant, the value of which never changes, for example, pi.

Data: numbers or measurements which are collected as a result of observations. They may be head counts (frequency data), as a number of individuals stating a preference for the Republican presidential candidate, or they may be scores, as on a psychological or educational test. Frequency data are also referred to as enumerative or categorical data.

Population or universe: a complete set of individuals, objects, or measurements having some common observable characteristic. Thus, all American citizens of voting age constitute a population.

Parameter: any characteristic of a *population* which is measurable, e.g., the proportion of registered Democrats among Americans of voting age. In this text we shall follow the practice of employing Greek letters (e.g. μ, σ) to represent population parameters.

Sample: a subset of a population or universe.

Statistic: a number resulting from the manipulation of raw data according to certain specified procedures. Commonly, we use a statistic which is calculated from a sample in order to estimate the population parameter, e.g., a sample of Americans of voting age is employed to estimate the proportion of Democrats in the entire population of voters. We shall employ italic letters (e.g. $\overline{X}$, s) to represent sample statistics. More will be said about the fascinating problem of sampling later in the text.

Example: Imagine an industrial firm engaged in the production of hardware for the space industry. Among its products are machine screws which must be maintained within fine tolerances with respect to width. As part of its quality control procedures, a number of screws are selected from the daily output and are carefully measured. These screws constitute the *sample.* The *variable* is the width of the screw. The *data* consist of the measurements of all screws collected in the sample. When the data are manipulated according to certain rules to yield certain summary statements, such as the "average" width of the screws, the resulting numerical value constitutes a *statistic.* The *population* to which we are interested in generalizing is the entire daily output of the plant. The "average" width of all the screws produced in a day constitutes a *parameter.* Note that it is highly unlikely that the parameter will ever be known, for to do so would require the measurement of every machine screw produced during the day. Since this is usually unfeasible for economic and other reasons, it is rare that an exhaustive study of populations is undertaken. Consequently, parameters are rarely known; but, as we shall see, they are commonly estimated from sample statistics.

Let us return to the two functions of statistical analysis for a closer look.

1.3 DESCRIPTIVE STATISTICS

When a behavioral scientist conducts a study, he characteristically collects a great deal of numerical information or data about the problem at hand. The data may take a variety of forms: frequency data (head counts of voters preferring various political candidates), or scale data (the weights of the contents of a popular breakfast cereal, or the I.Q. scores of a group of college students). In their original form, as collected, these data are usually a confusing hodge-podge of scores, frequency counts, etc. In performing the descriptive function, the statistician formulates rules and procedures for presentation of the data in a more usable and meaningful form. Thus the statistician states rules by which data may be represented graphically. He also formulates rules for calculating various *statistics* from masses of raw data.

Let us imagine that a behavioral scientist administered a number of measuring instruments (e.g., intelligence tests, personality inventories, aptitude tests) to a group of high school students. What are some things which he may do with the resulting measurements or scores to fulfill his descriptive functions?

1. He may rearrange the scores and group them in various ways in order to be able to see at a glance an overall picture of his data (Chapter 3, "Frequency Distributions and Graphing Techniques").

2. He may construct tables, graphs, and figures to permit visualization of the results (Section 3.3, "Graphing Techniques," in Chapter 3).

3. He may convert raw scores to other types of scores which are more useful for specific purposes. Thus, he may convert these scores into either percentile ranks, standard scores, or grades. Other types of conversions will also be described in the text (Chapter 4, "Percentiles," and Chapter 7, "The Standard Deviation And The Standard Normal Distribution").

4. He may calculate averages, to learn something about the typical performances of his subjects (Chapter 5, "Measures Of Central Tendency").

5. Employing the average as a reference point, he may describe the dispersion of scores about this central point. Statistics which quantify this dispersion are known as measures of variability or measures of dispersion (Chapter 6, "Measures Of Dispersion").

6. A relationship between two different measuring instruments may be obtained. The statistic for describing the extent of the relationship is referred to as a *correlation coefficient*. Such coefficients are extremely useful to the behavioral scientist. For example, he may wish to determine the relationship between intelligence and classroom grades, personality measures and aptitudes, or interests and personality measures. Once these relationships are established, the behavioral scientist may employ scores obtained from one measuring instrument to predict performance on another (Chapter 8, "Correlation," and Chapter 9, "Regression And Prediction").

1.4 INFERENTIAL STATISTICS

The behavioral scientist's task is not nearly over when he has completed his descriptive function. To the contrary, he is often nearer to the beginning than to the end of his task. The reason for this is obvious when we consider that the purpose of his research is often to explore hypotheses of a general nature rather than to simply compare limited samples.

Let us imagine that you are a behavioral scientist who is interested in determining the effects of a given drug upon the performance of a task involving psychomotor coordination. Consequently, you set up a study involving two conditions, experimental and control. You administer the drug to the experimental subjects at specified time periods before they undertake the criterion task. To rule out "placebo effects," you administer a pill containing inert ingredients to the control subjects. After all subjects have been tested, you perform your descriptive function. You find that "on the average" the experimental subjects did not perform as well as the controls. In other words, the arithmetic mean of the experimental group was lower than that of the control group. You then ask the question, "Can we conclude that the drug produced the difference between the two groups?" Or, more generally, "Can we assert that

the drug has an adverse effect upon the criterion task under investigation?" To answer these questions, it is not sufficient to rely solely upon *descriptive statistics.*

"After all," you reason, "even if the drug had *no effect*, it is highly improbable that the two group means would have been *identical. Some* difference would have been observed." The operation of uncontrolled variables (sometimes referred to rather imprecisely as "chance factors") is certain to produce some disparity between the group means. The critical question, from the point of view of inferential statistics, becomes: Is the difference great enough to rule out uncontrolled variation in the experiment as a sufficient explanation? Stated another way, if we were to repeat the experiment, would we be able to predict with confidence that the same differences (i.e., one mean greater than another) would systematically occur?

As soon as we raise these questions, we move into the fascinating area of statistical analysis which is known as *inductive* or *inferential* statistics. As you will see, much of the present text is devoted to procedures which the researcher employs to arrive at conclusions extending beyond the sample statistics themselves.

1.5 LYING WITH STATISTICS

A common misconception held by laymen is that statistics is merely a rather sophisticated method for fabricating lies or falsifying our descriptions of reality. The authors do not deny that some unscrupulous individuals employ statistics for just such purposes. However, such uses of statistics are anathema to the behavioral scientist who is dedicated to the establishment of truth. From time to time, references will be made to various techniques which are used for lying with statistics. However, the purpose is not to instruct you in these techniques but to make you aware of the various misuses of statistical analyses so that you do not inadvertently "tell a lie," and so that you may be aware when others do.

1.6 A WORD TO THE STUDENT

The study of statistics need not and should not become a series of progressive exercises in calculated tedium. If it is approached with the proper frame of mind, statistics can be one of the most exciting fields of study; it has applications in virtually all areas of human endeavor and cuts across countless fields of study. H. G. Wells, the 19th century prophet, remarked, "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write." Keep this thought constantly in mind. The course will be much more interesting and profitable to you if you develop the habit of "thinking statis-

tically." Constantly attempt to apply statistical concepts to all daily activities, no matter how routine. When you are stopped at an intersection which you cross frequently, note the time the traffic light remains red. Obtain some estimate of the length of the green cycle. If it is red three minutes and green two, you would expect that the chances are three in five that it will be red when you reach the intersection. Start collecting data. Do you find that it is red 60% of the time as expected? If not, why not? Perhaps you have unconsciously made some driving adjustments in order to change the statistical probabilities.

When you see statistical information being exhibited, develop a healthy attitude of skepticism. Ask pertinent questions. When a national magazine sends a shapely reporter to ten different diet doctors and receives an unneeded prescription from each, do not jump to the conclusion that "diet doctors" are frauds. Do not say, "After all, ten out of ten is a rather high proportion" and dismiss further inquiry at this point. Ask how the reporter obtained her sample. Was it at random or is it possible that the doctors were selected on the basis of prior information indicating they were rather careless in their professional practices? Question constantly, but reserve judgment until you have the answers.

Watch commercials on television; read newspaper advertisements. When the pitchman claims, "Dodoes are more effective," ask, "More effective than what? What is the evidence?"

If you make statistical thinking an everyday habit, you will not only find that the study of statistics becomes more interesting, but the world you live in will appear different and, perhaps, more interesting.

CHAPTER SUMMARY

In this chapter, we have distinguished between two definitions of statistical analysis, one stressing statistics as a collection of numerical facts and the other emphasizing statistics as a method or tool concerned with the collection, organization, and analysis of numerical facts. The second definition constitutes the subject matter of this text.

A distinction is made between two functions of the statistical method, descriptive and inferential statistical analyses. The former is concerned with the organization and presentation of data in a convenient, usable, and communicable form. The latter is addressed to the problem of making broader generalizations or inferences from sample data to populations.

A number of terms commonly employed in statistical analysis were defined.

Finally, it was pointed out that statistics is frequently employed for purposes of "telling lies." Such practices are inimical to the goal of establishing a factual basis for our conclusions and statistically-based decisions. However, you should be aware of the techniques for telling statistical lies so that you do

not inadvertently "tell one" yourself or fail to recognize one when someone else does.

New terms or concepts that have been introduced in a chapter will be listed at the end of each chapter. Some of these terms will be more precisely defined in other chapters and consequently may appear again.

Terms to Remember:

Statistical method
Descriptive statistics
Inferential or inductive statistics
Variable
Data
Population or universe
Parameter
Sample

EXERCISES

1. Which of the following most likely constitutes a *statistic*, a *parameter*, *data*, *inference from data*?
   a) A sample of 250 wage earners in Regent City yielded a *per capita* income of \$5650.
   b) The proportion of boys in a business arithmetic class is 0.58.
   c) Her "statistics" are 36, 12, 25.
   d) My tuition payment this year was \$1260.
   e) The number of people viewing Monday night's television special was 23,500,000.
   f) The birth rate in the United States increased by 5% over the previous month.
   g) The birth rate in Jerry Township increased by 3% over the previous month.
2. In your own words, describe what you understand the study of statistics to be.
3. A friend who is taking a course in accounting asks, "How does statistics differ from accounting? Both work with numbers and both are used in the field of business." What is your answer?
4. In the example cited in Section 1.2, let us imagine that, on a given day, the entire output of machine screws was measured. A summarizing statement such as, "The arithmetic average of the widths of all the screws was 0.23 mm." May we assume that we have established "truth" with respect to the widths of the screws for that particular day? What about the measurement problem?
5. While listening to the radio or viewing television, note the number of times that statistical data are cited during the commercials. How detailed are the citations? Is there possibly some "lying with statistics"?
6. Bring in newspaper examples citing recent survey or poll results. In how many articles is the method of sampling mentioned? Do the articles reveal where the financial support for the surveys came from? Why is this information important? Why is it so commonly not revealed?

7. List four populations. *Do not* include any that are defined by geographical boundaries.
8. Describe how you would select a random sample of:
   a) 50 registered Democrats in Phoenix, Arizona.
   b) 25 stocks from all those listed on the New York Stock Exchange.
   c) 30 utility tables from the population of those that are produced in a factory in Toledo, Ohio.
   d) 20 college students from all those enrolled at Fisher University.

Basic Mathematical Concepts 2

2.1 INTRODUCTION

"I'm not much good in math. How can I possibly pass statistics?" The authors have heard these words pass through the lips of countless undergraduate students. For many, this is probably a concern which legitimately stems from prior discouraging experiences with mathematics. A brief glance through the pages of this text may only serve to exacerbate this anxiety, since many of the formulas appear quite imposing to the novice and may seem impossible to master. Therefore, it is most important to set the record straight right at the beginning of the course.

You do not have to be a mathematical genius to master the statistical principles enumerated in this text. The amount of mathematical sophistication necessary for a firm grasp of the fundamentals of statistics is often exaggerated. As a matter of actual fact, statistics requires a good deal of arithmetic computation, sound logic, and a willingness to stay with a point until it is mastered. To paraphrase Carlyle, success in statistics is an infinite capacity for taking pains. Beyond these modest requirements, little is needed but the mastery of several algebraic and arithmetic procedures which most students learned early in their high school careers. In this chapter, we review the grammar of mathematical notations, discuss several types of numerical scales, and adopt certain conventions for the rounding of numbers.

For the student who wishes to brush up on his basic mathematics, Appendix I contains a review of all the math necessary to master this text.

2.2 THE GRAMMAR OF MATHEMATICAL NOTATION

Throughout the textbook, we shall be learning new mathematical symbols. For the most part, we shall define these symbols when they first appear. However, there are three notations which shall appear with such great regularity

that their separate treatment at this time is justified. These notations are $\sum$ (pronounced sigma), X, and N. However, while defining these symbols and showing their use, let's also review the grammar of mathematical notation.

It is not surprising to learn that many students become so involved in the forest of mathematical symbols, formulas, and operations, that they fail to realize that mathematics has a form of grammar which closely parallels the spoken language. Thus, mathematics has its nouns, adjectives, verbs, and adverbs.

Mathematical nouns. In mathematics, we commonly use symbols to stand for quantities. The notation we shall employ most commonly in statistics to represent quantity (or a score) is X, although we shall occasionally employ Y. In addition, X and Y are employed to identify variables; for example, if weight and height were two variables in a study, X might be used to represent weight and Y to represent height. Another frequently used "noun" is the symbol N which represents the number of scores or quantities with which we are dealing. Thus, if we have ten quantities,

$$N = 10.$$

Mathematical adjectives. When we want to modify a mathematical noun, to identify it more precisely, we commonly employ subscripts. Thus if we have a series of scores or quantities, we may represent them as X_1, X_2, X_3, X_4, etc. We shall also frequently encounter X_i, in which the subscript may take on any value that we desire.

Mathematical verbs. Notations which direct the reader to do something have the same characteristics as verbs in the spoken language. One of the most important "verbs" is the symbol already alluded to as $\sum$. This notation directs us to sum all quantities or scores following the symbol. Thus,

$$\sum(X_1, X_2, X_3, X_4, X_5)$$

indicates that we should add together all these quantities from X_1 through X_5. Other "verbs" we shall encounter frequently are $\sqrt{\ }$, directing us to find the square root, and exponents (X^a), which tell us to raise a quantity to the indicated power. In mathematics, mathematical verbs are commonly referred to as operators.

Mathematical adverbs. These are notations which, as in spoken language, modify the verbs. We shall frequently find that the summation signs are modified by adverbial notations. Let us imagine that we want to indicate that the following quantities are to be added:

$$X_1 + X_2 + X_3 + X_4 + X_5 + \cdots + X_N.$$

Symbolically, we would represent these operations as follows:

$$\sum_{i=1}^{N} X_i.$$

The notations above and below the summation sign indicate that i takes on the successive values from 1, 2, 3, 4, 5 up to N. Stated verbally, the notation reads: We should sum all quantities of X starting with $i = 1$ (that is, X_1) and proceeding through to $i = N$ (that is, X_N).

Sometimes this form of notation may direct us to add only selected quantities; thus,

$$\sum_{i=2}^{5} X_i = X_2 + X_3 + X_4 + X_5.$$

2.3 SUMMATION RULES

The summation sign is one of the most frequently occurring operators in statistics. Let us summarize a few of the rules governing the use of the summation sign.

Imagine a sample in which $N = 3$ and $X_1 = 3$, $X_2 = 4$, and $X_3 = 6$. The sum of the three values of the variable may be shown by

$$\begin{aligned}\sum_{i=1}^{N} X_i &= X_1 + X_2 + X_3\\ &= 3 + 4 + 6.\end{aligned}$$

Let a be a constant. To show the sum of the values of a variable when a constant has been added to each,

$$\begin{aligned}\sum_{i=1}^{N} (X_i + a) &= (3 + a) + (4 + a) + (6 + a)\\ &= 3 + 4 + 6 + (a + a + a)\\ &= 13 + 3a.\end{aligned}$$

Thus

$$\sum_{i=1}^{N} (X_i + a) = \sum_{i=1}^{N} X_i + Na.$$

Generalization: *The sum of the values of a variable plus a constant is equal to the sum of the values of the variable plus N times that constant.*

To show the sum of the values of a variable when a constant has been subtracted from each,

$$\begin{aligned}\sum_{i=1}^{N}(X_i - a) &= (3-a)+(4-a)+(6-a)\\ &= 3+4+6-(a+a+a)\\ &= 13-3a.\end{aligned}$$

Thus

$$\sum_{i=1}^{N}(X_i - a) = \sum_{i=1}^{N} X_i - Na.$$

Generalization: *The sum of the values of a variable when a constant has been subtracted from each is equal to the sum of the values of the variable minus N times the constant.*

Example: $\sum_{i=1}^{N}(X_i - \overline{X}) = \sum_{i=1}^{N} X_i - N\overline{X}.$

2.4 TYPES OF NUMBERS

Cultural anthropologists, psychologists, and sociologists have repeatedly called attention to man's tendency to explore and understand the world that is remote from his primary experiences, long before he has investigated that which is closest to him. Thus, while man was probing distant stars and describing with striking accuracy their apparent movements and their interrelationships, he virtually ignored the very substance which gave him life: air which he inhales and exhales over four hundred million times a year. In the authors' experience, a similar pattern exists in relation to the student's experience with numbers and his concepts of them. In our very quantitatively oriented western civilization, the child employs and manipulates numbers long before he is expected to calculate the batting averages of the latest baseball hero. Nevertheless, ask him to define a number, or to describe the ways in which numbers are employed, and you will likely be met with expressions of consternation and bewilderment. "I have never thought about it before," he will frequently reply. After a few minutes of soul searching and deliberation, he will probably reply something to the effect that numbers are symbols which denote amounts of things which can be added, subtracted, multiplied, and divided. These are all familiar arithmetic concepts, but do they exhaust all possible uses of numbers? At the risk of reducing our student to utter confusion, you may ask: "Is the symbol 7 on a baseball player's uniform such a number? What about your home address? Channel 2 on your television set? Do these numbers indicate amounts of things? Can they reasonably be added, subtracted, multiplied, or divided? Can you multiply the number on any football player's back by any other number and obtain a meaningful value?" A careful analysis of our use of numbers in every-

day life reveals a very interesting fact: most of the numbers we employ do not have the arithmetical properties we usually ascribe to them. For this reason, we prefer to differentiate between two terms, "numbers" and "numerals." Numerals refer to symbols such as Y, 10, IX. Numbers are specific types of numerals which bear fixed relationships to other numerals. Thus, two numerals such as 4 and 5 are numbers if, and only if, they can be meaningfully added, multiplied, subtracted, and divided. From this point on the terms "number" and "numeral" will be differentiated on this basis. The list of such numerals is large. A few examples in addition to those enumerated above are: the serial number on a home appliance, a Zipcode number, a telephone number, a home address, an automobile registration number, and the numbers on a book in the library.

The important point is that numerals are used in a variety of ways to achieve many different ends. Much of the time, these ends do not include the representation of an amount or a quantity. In fact, there are three fundamentally different ways in which numerals are used.

1. To name (nominal numerals)
2. To represent position in a series (ordinal numerals)
3. To represent quantity (numbers)

2.5 TYPES OF SCALES

The fundamental requirements of observation and measurement are acknowledged by all the physical and social sciences as well as by any modern-day corporation interested in improving its competitive position. The things that we observe are often referred to as *variables* or *variates*. For example, if we are studying the price of stocks on the New York exchange, our variable is price. Any particular observation is called the *value of the variable*, or a score.

2.5.1 Nominal Scales

It is probable that when the majority of people think about measurement, they conjure up mental images of wild-eyed men in white suits manipulating costly and incredibly complex instruments in order to obtain precise measures of the variable that they are studying. Actually, however, not all measurements are this precise or this quantitative. If we were to study the sex of the offspring of female rats which had been subjected to atomic radiation during pregnancy, sex would be the variable that we would observe. There are only two possible values of this variable: male and female (barring an unforseen mutation which produced a third sex!). Our data would consist of the number of observations in each of these two classes. Note that we do not think of this variable as representing an ordered series of values, such as height, weight, speed, etc.

An organism which is female does not have any more of the variable, sex, than one which is male, in spite of what Hollywood tries to tell us.

When is a woman a woman? Assignment of individuals or objects to classes is not always as clear-cut as it might first appear, since the properties defining the class are not always universally agreed upon. The recent controversy over the sex of females in international athletic competition is a case in point. Since all female athletes are required to submit to a series of sex tests prior to international competition, a number of renowned "female" athletes from behind the "iron curtain" have disappeared from the international scene. One is led to suspect that they might not have passed the physical, for one reason or another. The most fascinating and controversial case involves the great Polish athlete Ewa Klobkowska who passed the physical examination but was later disqualified when the study of her chromosomes revealed the presence of "masculine" Y-chromosomes. Is "she" female or male? An interesting outgrowth of the controversy has been the demand by some female athletes—perhaps with tongue-in-cheek—that the women doctors charged with examining them submit themselves to a prior physical examination to verify their "true" sex.

Observations of unordered variables constitute a very low level of measurement and are referred to as a nominal scale of measurement. We may assign numerical values to represent the various classes in a nominal scale but these numbers have no quantitative properties. They serve to identify the class.

The data employed with nominal scales consist of frequency counts or tabulations of the number of occurrences in each class of the variable under study. In the aforementioned radiation study, our frequency counts of male and female progeny would comprise our data. Such data are often referred to interchangeably as *frequency data*, *enumerative data*, *attribute data*, or *categorical data*. The only mathematical relationships germane to nominal scales are those of equivalence ($=$) or of nonequivalence ($\neq$).

2.5.2 Ordinal Scales

When we move into the next higher level of measurement, we encounter variables in which the classes *do* represent an ordered series of relationships. Thus, the classes in ordinal scales are not only different from one another (the characteristic defining nominal scales) but they stand in some kind of *relation* to one another. More specifically, the relationships are expressed in terms of the algebra of inequalities: a is less than b ($a < b$) or a is greater than b ($a > b$). Thus the relationships encountered are: greater, faster, more intelligent, more mature, more prestigious, more disturbed, etc. The numerals employed in connection with ordinal scales are nonquantitative. They indicate only position in an ordered series and not "how much" of a difference exists between successive positions on the scale.

Examples of ordinal scaling include: rank ordering of baseball players according to their "value to the team," rank ordering of laboratory rats according to their "speed" in learning to run a maze, rank ordering of potential candidates for political office according to their "popularity" with people, and rank ordering of officer candidates in terms of their "leadership" qualities. Note that the ranks are assigned according to the ordering of individuals within the class. Thus, the most popular candidate may receive the rank of 1, the next popular may receive the rank of 2, and so on, down to the least popular candidate. It does not, in fact, make any difference whether or not we give the most popular candidate the highest numerical rank or the lowest, *so long as we are consistent in placing the individuals accurately with respect to their relative position in the ordered series.* By popular usage, however, the lower numerical ranks (1st, 2nd, 3rd) are usually assigned to those "highest" on the scale. Thus, the winning horse receives the rank of "first" in a horse race; the pennant winner is "first" in its respective league; the rat requiring fewest trials to run a maze is "first" in its running performance. The fact that we are not completely consistent in our ranking procedures is illustrated by such popular expressions as "first-class idiot" and "first-class scoundrel."

2.5.3 Interval and Ratio Scales

Finally, the highest level of measurement in science is achieved with scales employing numbers (*interval* and *ratio* scales). The numerical values associated with these scales are truly quantitative and therefore permit the use of arithmetic operations such as adding, subtracting, multiplying, and dividing. In interval and ratio scales equal differences between points on any part of the scale are equal. Thus the difference between 4 feet and 2 feet is the same as the difference between 9231 and 9229 feet.

In ordinal scaling, as you will recall, we could not claim that the difference between the first and second horses in a race was the same as the difference between the second and third horses.

There are two types of scales based upon numbers: interval and ratio. The only difference between the two scales stems from the fact that the interval scale employs an arbitrary zero point, whereas the ratio scale employs a true zero point. Consequently, only the ratio scale permits us to make statements concerning the ratios of numbers in the scale; e.g., 4 feet are to 2 feet as 2 feet are to 1 foot. A good example of the difference between an interval and a ratio scale is height as measured from a table top (interval) vs. height as measured from the floor. The difference between these scales can be further clarified by examining a well-known interval scale, e.g., the centigrade scale of temperature. Incidentally, the Fahrenheit scale of temperature is also an interval scale.

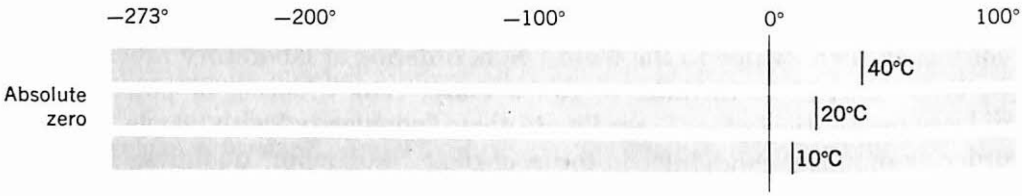


Fig. 2.1 Relationships of various points on a centigrade scale to absolute zero.

The zero point on the centigrade scale does not represent the complete absence of heat. In fact, it is merely the point at which water freezes at sea level and it has, therefore, an arbitrary zero point. Actually, the true zero point is known as absolute zero which is approximately $-273°$ centigrade. Now, if we were to say that 40°C is to 20°C as 20° C is to 10°C, it would appear that we were making a correct statement. Actually, we are completely wrong since 40°C really represents $273° + 40°$ of heat; 20°C represents $273° + 20°$ of heat; and 10°C represents $273° + 10°$ of heat. The ratio 313:293 as 293:283, clearly does not hold. These facts may be better appreciated graphically. In Fig. 2.1, we have represented all three temperature readings as distances from the true zero point which is -273°C. From this graph it is seen that the distance from -273°C to 40°C is not twice as long as the distance from -273°C to 20°C. Thus, 40°C is not twice as warm as 20°C, and the ratio 40°C is to 20°C as 20°C is to 10°C does not hold.

Apart from the difference in the nature of the zero point, interval and ratio scales have the same properties and will be treated alike throughout the text.

It should be clear that one of the most sought-after goals of the behavioral scientist is to achieve measurements which are at least interval in nature. Unfortunately, behavioral science has met with little success along these lines. Indeed, there are certain epistemological considerations which make the authors doubt that we will ever achieve interval scaling for many of the types of dimensions which behavioral scientists measure. In several instances where interval scaling is claimed, it is by virtue of certain assumptions which the claimant is willing to make. For example, some specialists in scaling procedures will assume that "equal-appearing intervals are equal" as the basis for claiming cardinality. The approximation of such scales to cardinal measurement is, of course, only as good as the validity of the assumption.

2.6 CONTINUOUS AND DISCONTINUOUS SCALES

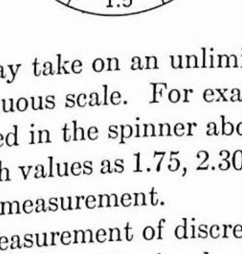
Let us imagine that you are given the problem of trying to determine the number of children per American family. Your scale of measurement would start with zero (no children) and would proceed, by *increments of one* to perhaps

fifteen or twenty. Note, that in moving from one value on the scale to the next, we proceed by *whole numbers* rather than by fractional amounts. Thus a family has either 0, 1, 2, or more children. In spite of the statistical abstraction that the American family averages two and a fraction children, the authors do not know a single couple that has achieved this marvelous state of family planning.

Such scales, in which the variable can take on a finite number of values, are referred to as *discontinuous* or *discrete* scales, and they have equality of *counting units* as their basic characteristic. Thus, if we are studying the number of children in a family, each child is equal with respect to providing one counting unit. Such scales involve cardinality insofar as they permit arithmetic operations such as adding, subtracting, multiplying, and dividing. Thus we can say that a family with four children has twice as many children as one with two children. Observations of discrete variables are always exact so long as the counting procedures are accurate. Examples of discontinuous variables are: the number of hats sold in a department store during the month of January, the number of white blood cells counted in one square centimeter, the number of alpha particles observed in a second, etc.

You should not assume from the above discussion that discrete scales necessarily involve *only* whole numbers. A spinner, such as the one shown in Fig. 2.2, would be discrete, yet the seven values of the variable proceed by half units. The important point, in this example, is that the variable cannot take on values *between* 0 and 0.5, or 0.5 and 1, etc.

Fig. 2.2 A spinner illustrating a discrete scale in which the variable changes by half units.



In contrast, a scale in which the variable may take on an unlimited number of intermediate values is referred to as a continuous scale. For example, let us take the same range of values which is illustrated in the spinner above (0 to 3). If it were possible for this variable to take on such values as 1.75, 2.304, etc., then we would be dealing with a continuous scale of measurement.

It is important to note that, although our measurement of discrete variables is always exact, our measurement of continuous variables is always approximate. If we are measuring the height of American males, for example, any particular measurement is inexact because it is always possible to imagine a measuring stick which would provide greater accuracy. Thus, if we reported the height of a man to be 68 inches, we would mean 68 inches give or take one half an inch. If our scale is accurate to the nearest tenth, we can always imagine another scale providing greater accuracy, say, to the nearest hundredth or

thousandth of an inch. The basic characteristic of continuous scales then, is equality of *measuring units*. Thus, if we are measuring in inches, one inch is always the same throughout the scale. Examples of continuous variables are length, velocity, time, weight, etc.

2.6.1 Continuous Variables, Errors of Measurement, and "True Limits" of Numbers

In our preceding discussion, we pointed out that continuously distributed variables can take on an unlimited number of intermediate values. Therefore, we can never specify the exact value for any particular measurement, since it is possible that a more sensitive measuring instrument can increase the accuracy of our measurements a little more. For this reason, we stated that numerical values of continuously distributed variables are always approximate. However, it is possible to specify the limits within which the true value falls; e.g., the true limits of a value of a continuous variable are equal to that number plus or minus one half of the unit of measurement. Let us look at a few examples. You have a bathroom scale, which is calibrated in terms of pounds. When you step on the scale the pointer will usually be a little above or below a pound marker. However, you report your weight to the nearest pound. Thus, if the pointer were approximately three quarters of the distance between 212 pounds and 213 pounds, you would report your weight as 213 pounds. It would be understood that the "true" limit of your weight, assuming an accurate scale, falls between 212.5 pounds and 213.5 pounds. If, on the other hand, you are measuring the weight of whales, you would probably have a fairly gross unit of measurement, say 100 pounds. Thus, if you reported the weight of a whale at 32,000 pounds, you would mean that the whale weighed between 31,950 pounds and 32,050 pounds. If the scale were calibrated in terms of 1000 pounds, the true limits of the whale's weight would be between 31,500 pounds and 32,500 pounds.

2.7 ROUNDING

Let us imagine that we have obtained some data which, in the course of conducting our statistical analysis, require that we divide one number into another. There will be innumerable occasions in this course when we shall be required to perform this arithmetic operation. In most cases, the answer will be a value which extends to an endless number of decimal places. For example, if we were to express the fraction $\frac{1}{3}$ in decimal form, the result would be 0.33333+. It is obvious that we cannot extend this series of numbers *ad infinitum*. We must terminate at some point and assign a value to the last number in the series which best reflects the remainder. When we do this, two types of problems will arise:

1. To how many decimal places do we carry the final answer?
2. How do we decide on the last number in the series?

The answer to the first question is usually given in terms of the number of significant figures. However, there are many good reasons for not following the mathematical stricture to the letter. For simplicity and convenience, we shall adopt the following policy with respect to rounding:

After every operation, carry to three and round to two more places than were in the original data.

Thus, if the original data were in whole-numbered units, we would carry our answer to the third decimal place and round to the second decimal. If in tenths, we would round to the third decimal, and so forth.

Once we have decided the number of places to carry our final figures, we are still left with the problem of representing the last digit. Fortunately, the rule governing the determination of the last digit is perfectly simple and explicit. If the remainder beyond that digit is greater than 5, increase that digit to the next higher number. If the remainder beyond that digit is less than 5, allow that digit to remain as it is. Let's look at a few illustrations. In each case, we shall round to the second decimal place:

6.546	becomes	6.55,
6.543	becomes	6.54,
1.967	becomes	1.97,
1.534	becomes	1.53.

You may ask, "In the above illustrations, what happens if the digit at the third decimal place is five?"

You should first determine whether or not the digit is exactly 5. If it is 5 plus the slightest remainder, the above rule holds and you must add one to the digit at the second decimal place. If it is almost, but not quite 5, the digit at the second decimal place remains the same. If it is *exactly 5, with no remainder,* then an arbitrary convention which is accepted universally by mathematicians applies: Round the digit at the second decimal place to the *nearest even* number. If this digit is already even, then it is not changed. If it is odd, then *add* one to this digit to make it even. Let's look at several illustrations in which we round to the second decimal place:

6.545001	becomes	6.55. Why?
6.545000	becomes	6.54. Why?
1.9652	becomes	1.97. Why?
0.00500	becomes	0.00. Why?
0.01500	becomes	0.02. Why?
16.89501	becomes	16.90. Why?

2.8 FREQUENCY, PROPORTION, AND PERCENTAGE

Throughout the text we shall be dealing with data in terms of frequency, proportion, and percentage. You should be aware of the relationship among these three terms.

If there are 300 women in a class of 1200 students, the frequency of women is 300. The proportion (p) of women to the total is 300/1200 or, expressed decimally, $p = 0.25$. Finally, to express this proportion as a percentage, you simply multiply by 100, i.e., the percentage of women is 25% of the total:

$$\frac{\text{frequency of women}}{\text{total frequency}} = \text{proportion of women} = \frac{\text{percentage of women}}{100}.$$

CHAPTER SUMMARY

In this chapter, we pointed out that advanced knowledge of mathematics is not a prerequisite for success in this course. A sound background in high school mathematics plus steady application to assignments should be sufficient to permit mastery of the fundamental concepts put forth in this text.

To aid the student who may not have had recent contact with mathematics, we have attempted to review some of the basic concepts of mathematics. Included in this review are: (1) the grammar of mathematical notations, (2) types of numbers, (3) types of numerical scales, (4) continuous and discontinuous scales, (5) rounding and (6) ratios, frequencies, proportions and percentages. Students requiring a more thorough review of mathematics may refer to Appendix I.

Terms to Remember:

Σ
Nominal numeral
Ordinal numeral
Cardinal number
Variable or Variate
Value of variable
Nominal scale
Frequency data
Enumerative data
Attributive data
Categorical data
Ordinal scale
Interval scale
Ratio scale
Discontinuous scales
Continuous scales
Counting units
Measuring units
True limit of a number
Rounding
Frequency
Proportion
Percentage

EXERCISES

The following exercises are based on this chapter and Appendix I.

1. Solve the following equations for X.
   a) $a + X = b - c$ b) $2X + a = c$ c) $aX + bY = c$
   d) $\frac{a}{X} = \frac{b}{c}$ e) $\frac{b}{c/(X/a)} = Y$ f) $\frac{a + X}{Y/(b/c)} = c$
   g) $\frac{a}{b + X} = c$ h) $(X^3)(X^4)(X^2) = Y$ i) $\frac{X^5}{X^8} = aX^2$
2. Substitute in the following equations and solve for X.
   a) $aX + bY = c$, in which $a = 3$, $b = 2$, $Y = 8$, $c = 25$
   b) $a^5 - a^3 = X$, in which $a = 2$
   c) $(a^2)(b^3) = X$, in which $a = \frac{1}{6}$, $b = \frac{5}{6}$
   d) $\frac{a^5}{a^2} = X$, in which $a = 3$
   e) $a^3 - \frac{b}{c} = X$, in which $a = 2$, $b = \frac{2}{3}$, $c = \frac{3}{4}$

3. Round the following numbers to the second decimal place.
   a) 99.99500 b) 46.40501
   c) 2.96500 d) 0.00501
   e) 16.46500 f) 1.05499
   g) 86.2139 h) 10.0050
4. Determine the answers to the following problems to as many places as is standard procedure.
   a) 0.275 times 0.111 b) 0.3811 times 0.2222
   c) 0.999 times 0.121 d) 150 divided by 400
   e) 0.1 divided by 0.9 f) 0.006 divided by 0.007
5. Determine the proportion and percentage of the following.
   a) Male students in your statistics class.
   b) Number of face cards in a conventional 52-card deck.
   c) 25 items correct on a quiz consisting of 33 items.
   d) 4652 voters out of a total registration of 9686.
6. Determine the value of the following expressions in which $X_1 = 4$, $X_2 = 5$, $X_3 = 7$, $X_4 = 9$, $X_5 = 10$, $X_6 = 11$, $X_7 = 14$.
   a) $\sum_{i=1}^{4} X_i =$ b) $\sum_{i=1}^{7} X_i =$ c) $\sum_{i=3}^{6} X_i =$
   d) $\sum_{i=2}^{5} X_i =$ e) $\sum_{i=1}^{N} X_i =$ f) $\sum_{i=4}^{N} X_i =$
7. Express the following in summation notation.
   a) $X_1 + X_2 + X_3$ b) $X_1 + X_2 + \cdots + X_N$
   c) $X_3^2 + X_4^2 + X_5^2 + X_6^2$ d) $X_4^2 + X_5^2 + \cdots + X_N^2$

8. The answers to the following questionnaire items are based on what scale of measurement?
   a) What is your height?
   b) What is your weight?
   c) What is your occupation?
   d) How does this course compare with others you have taken?
   e) What is your name?
9. In the following examples, identify the scale of measurement, and determine whether the italicized variable is continuous or discontinuous.
   a) *Distance* traveled from home to school.
   b) Number of infants born at *varying times of the day.*
   c) Number of votes compiled by each of *three candidates* for a political office.
   d) Number of servicemen at varying *ranks in the U.S. Army.*
10. Determine the square roots of the following numbers to two decimal places.
    a) 160 b) 16 c) 1.60 d) 0.16 e) 0.016
11. State the true limits of the following numbers.
    a) 0 b) 0.5 c) 1.0 d) 0.49 e) −5 f) −4.5
12. Using the values of X_i given in Problem 6 above, show that

$$\sum_{i=1}^{N} X_i^2 \neq \left(\sum_{i=1}^{N} X_i\right)^2.$$

13. Show that

$$\sum_{i=1}^{N} X_i Y_i \neq \sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i$$

is a true statement by giving an example.

14. Discuss the special characteristics of each scale of measurement. Give two examples of each type of scale.
15. Which of the following represent continuous scales and which represent discrete scales of measurement?
    a) The number of light bulbs sold each day in a hardware store.
    b) The monthly income of graduate students.
    c) The temperatures recorded every two hours in the meat department of a supermarket.
    d) The weights of pigeons recorded every 24 hours in an experimental laboratory.
    e) The lengths of newborns recorded at a hospital nursery.
    f) The number of textbooks found in the college bookstore.
16. Using the figures shown in the table below, answer the following questions:
    a) Of all the students majoring in each academic area, what percentage is female?
    b) Considering only the males, what percentages are found in each academic area?
    c) Considering only the females, what percentages are found in each academic area?

d) Of all students majoring in the five areas, what percentage is male? What percentage is female?

The number of students, by sex, majoring in each of five academic areas:

	Male	Female
Business administration	400	100
Education	50	150
Humanities	150	200
Science	250	100
Social science	200	200

17. A large discount house advertises, "Prices reduced below cost." What does this claim mean to you? Ask your friends to interpret the claim. Are the interpretations all in agreement?

18. Refer to Table 3.4, and answer the following. If all the scores were placed in a hat and selected at random, what is the likelihood that:
    a) The score would be in the interval 145–149?
    b) The score would be between 130 and 138?
    c) The score would be 94 or below?
    d) The score would be 145 or above?

19. Indicate which of the following variables represent discrete or continuous series.
    a) The time it takes you to complete these problems.
    b) The number of newspapers sold in a given city on December 19, 1969.
    c) The amount of change in weight of 5 women during a period of 4 weeks.
    d) The number of home runs hit by 10 pitchers, selected at random, during the 1970 baseball season.
    e) The number of stocks on the New York Stock Exchange that increased in selling price on January 3, 1970.

20. Following is a list showing the number of births in the United States (expressed in thousands) between 1940 and 1960. (Source: *The World Almanac*, 1967 (adapted). Published by Newspaper Enterprise Association Inc., N.Y., p. 684.)

Year	Males	Females
1940	1212	1149
1945	1405	1331
1950	1824	1731
1955	2074	1974
1960	2180	2078

Calculate the percentage of males and females for each year.

Frequency Distributions and Graphing Techniques 3

3.1 GROUPING OF DATA

Let us imagine that you have just accepted a position as curriculum director in a large senior high school. Your responsibility is to develop curricula which are in agreement with the needs and the motivations of the students, and at the same time, provide maximum challenges to their intellectual capacities. Now, it is beyond the scope of the text to provide a solution to this very complex and provoking problem. However, it is clear that no steps toward a solution can be initiated without some assessment of the intellectual capacities of the student body. Accordingly, you go to the Guidance Office and "pull out" *at random* (i.e., in such a way that every member of the population shares an equal chance of being selected) 110 student dockets containing a wealth of personal and scholastic information. Since your present interest is to assess intellectual ability, you focus your attention upon the entry labeled, "I.Q. estimate." You write these down on a piece of paper, with the results listed in Table 3.1.

As you mull over these figures, it becomes obvious to you that you cannot "make heads or tails" out of them unless you organize them in some systematic

Table 3.1

I.Q. Scores of 110 high school students selected at random

154	131	122	100	113	119	121	128	112	93
133	119	115	117	110	104	125	85	120	135
116	103	103	121	109	147	103	113	107	98
128	93	90	105	118	134	89	143	108	142
85	108	108	136	115	117	110	80	111	127
100	100	114	123	126	119	122	102	100	106
105	111	127	108	106	91	123	132	97	110
150	130	87	89	108	137	124	96	111	101
118	104	127	94	115	101	125	129	131	110
97	135	108	139	133	107	115	83	109	116
110	113	112	82	114	112	113	142	145	123

Table 3.2
Frequency distribution of I.Q. scores of 110 high school students selected at random

<table>
<tr><th>X</th><th>f</th><th>X</th><th>f</th><th>X</th><th>f</th><th>X</th><th>f</th></tr>
<tr><td>154</td><td>|</td><td>135</td><td>||</td><td>116</td><td>||</td><td>97</td><td>||</td></tr>
<tr><td>153</td><td></td><td>134</td><td>|</td><td>115</td><td>||||</td><td>96</td><td>|</td></tr>
<tr><td>152</td><td></td><td>133</td><td>||</td><td>114</td><td>||</td><td>95</td><td></td></tr>
<tr><td>151</td><td></td><td>132</td><td>|</td><td>113</td><td>||||</td><td>94</td><td>|</td></tr>
<tr><td>150</td><td>|</td><td>131</td><td>||</td><td>112</td><td>|||</td><td>93</td><td>||</td></tr>
<tr><td>149</td><td></td><td>130</td><td>|</td><td>111</td><td>|||</td><td>92</td><td></td></tr>
<tr><td>148</td><td></td><td>129</td><td>|</td><td>110</td><td>|||||</td><td>91</td><td>|</td></tr>
<tr><td>147</td><td>|</td><td>128</td><td>||</td><td>109</td><td>||</td><td>90</td><td>|</td></tr>
<tr><td>146</td><td></td><td>127</td><td>|||</td><td>108</td><td>||||||</td><td>89</td><td>||</td></tr>
<tr><td>145</td><td>|</td><td>126</td><td>|</td><td>107</td><td>||</td><td>88</td><td></td></tr>
<tr><td>144</td><td></td><td>125</td><td>||</td><td>106</td><td>||</td><td>87</td><td>|</td></tr>
<tr><td>143</td><td>|</td><td>124</td><td>|</td><td>105</td><td>||</td><td>86</td><td></td></tr>
<tr><td>142</td><td>||</td><td>123</td><td>|||</td><td>104</td><td>||</td><td>85</td><td>||</td></tr>
<tr><td>141</td><td></td><td>122</td><td>||</td><td>103</td><td>|||</td><td>84</td><td></td></tr>
<tr><td>140</td><td></td><td>121</td><td>||</td><td>102</td><td>|</td><td>83</td><td>|</td></tr>
<tr><td>139</td><td>|</td><td>120</td><td>|</td><td>101</td><td>||</td><td>82</td><td>|</td></tr>
<tr><td>138</td><td></td><td>119</td><td>|||</td><td>100</td><td>||||</td><td>81</td><td></td></tr>
<tr><td>137</td><td>|</td><td>118</td><td>||</td><td>99</td><td></td><td>80</td><td>|</td></tr>
<tr><td>136</td><td>|</td><td>117</td><td>||</td><td>98</td><td>|</td><td></td><td></td></tr>
</table>

fashion. It occurs to you to list all the scores from highest to lowest and then place a slash mark alongside of each score every time it occurs (Table 3.2). The number of slash marks, then, represents the frequency of occurrence of each score.

When you have done this, you have constructed an ungrouped frequency distribution of scores. Note that, in the present example, the scores are widely spread out, a number of scores have a frequency of zero, and there is no "visually" clear indication of central tendency. Under these circumstances, it is customary for most researchers to "group" the scores into what is referred to as *class intervals* and then obtain a frequency distribution of "grouped scores."

3.1.1 Grouping into Class Intervals

Grouping into class intervals involves a sort of "collapsing the scale" in which we assign scores to *mutually exclusive** classes in which the classes are defined in terms of the grouping intervals employed. The reasons for grouping are

* We refer to the classes as *mutually exclusive* because it is impossible for a subject's score to belong to more than one class.

twofold: (1) It is uneconomical and unwieldy to deal with a large number of cases spread out over many scores unless automatic calculators are available. (2) Some of the scores have such low frequency counts associated with them that it is not warranted to maintain these scores as separate and distinct entities.

On the negative side is, of course, the fact that grouping inevitably results in the loss of information. For example, individual scores lose their identity when we group into class intervals and some small errors in statistics based upon grouped scores are unavoidable.

The question now becomes, "On what basis do we decide upon the grouping intervals which we will employ?" Obviously, the interval selected must not be so gross that we lose the discrimination provided by our original measurement. For example, if we were to divide the previously collected I.Q. scores into two classes; those below 100 and those 100 and above, practically all the information inherent in the original scores would be lost. On the other hand, the class intervals should not be so fine that the purposes served by grouping are defeated. In answer to our question, there is, unfortunately, no general prescription which can be applied to all data. Much of the time the choice of the number of class intervals must represent a judgment based upon a consideration of the relative effects of grouping upon discriminability and presentational economy. However, it is generally agreed that most data in the behavioral sciences can be accommodated by 10 to 20 class intervals. For uniformity, we shall aim for approximately 15 class intervals for the data that we shall discuss in this text.

Having decided upon the number of class intervals that is appropriate for a set of data, the procedures for assigning scores to class intervals are quite straightforward. Although one of several different techniques may be used, we shall employ only one for the sake of consistency. The procedures to be employed are as follows:

Step 1. Find the difference between the highest and the lowest score values contained in the original data. Add 1 to obtain the total number of scores or potential scores. In the present example, this result is $(154–80) + 1 = 75$.

Step 2. Divide this figure by 15 to obtain the number of scores or potential scores in each interval. If the resulting value is not a whole number, and it usually is not, the authors prefer to round to the nearest odd number so that a whole number will be at the middle of the class interval. However, this practice is far from universal and you would not be wrong if you rounded to the *nearest number*. In the present example, the number of scores for each class interval is $\frac{75}{15}$, or 5. We shall designate the class interval by the symbol i. In the example, $i = 5$.

Step 3. Take the lowest score in the original data as the minimum value in the lowest class interval. Add to this $i - 1$ to obtain the maximum score of the lowest class interval. Thus, the lowest class interval of the data on hand is $80 - 84$.

Step 4. The next higher class interval begins at the integer following the maximum score of the lower class interval. In the present example, the next integer is 85. Follow the same steps as in step 3 to obtain the maximum score of the second class interval. Follow these procedures for each successive higher class interval until all the scores are included in their appropriate class intervals.

Step 5. Assign each obtained score to the class interval within which it is included. The *grouped frequency distribution* appearing in Table 3.3 was obtained by employing the above procedures.

Table 3.3

Grouped frequency distribution of I.Q. scores based upon data appearing in Table 3.2

Class interval	f	Class interval	f
150–154	2	110–114	17
145–149	2	105–109	14
140–144	3	100–104	12
135–139	5	95–99	4
130–134	7	90–94	5
125–129	9	85–89	5
120–124	9	80–84	3
115–119	13		

$N = 110$

You will note that by grouping we may obtain an immediate "picture" of the distribution of I.Q. scores among our high school students. For example, we note that there is a clustering of frequencies in the class intervals between the scores of 100 and 119. It is also apparent that the number of scores in the extremes tends to dwindle off. Thus, we have achieved one of our objectives in grouping, to provide an economical and manageable array of scores.

3.1.2 The True Limits of a Class Interval

In our prior discussion of the "true limits" of a number, (Section 2.6.1) we pointed out that the "true" value of a number is equal to its apparent value plus and minus one-half of the unit of measurement. Of course, the same is true of these values even after they have been grouped into class intervals. Thus, although we write the limits of the lowest class interval as 80–84, the true limits of the interval are 79.5–84.5 (i.e., the lower real limit of 80 and the upper real limit of 84, respectively).

It is important to keep in mind that the true limits of a class interval are not the same as the *apparent limits*. When calculating the median and percentile ranks for grouped data, we shall make use of the *true limits* of the class interval.

3.2 CUMULATIVE FREQUENCY AND CUMULATIVE PERCENTAGE DISTRIBUTIONS

It is often desirable to arrange the data from a frequency distribution into a cumulative frequency distribution. Besides aiding in the interpretation of the frequency distribution, a cumulative frequency distribution is of great value in obtaining the median and the various percentile ranks of scores, as we shall see in Chapter 4.

The cumulative frequency distribution is obtained in a very simple and straightforward manner. Let us attend to the data in Table 3.4.

The entries in the frequency distribution indicate the number of high school students falling within each of the class intervals. Each entry within the cumulative frequency distribution indicates the number of all cases or frequencies *below the upper real limit* of that interval. Thus, in the third class interval from the bottom in Table 3.4, the entry "13" in the cumulative frequency distribution indicates that a total of 13 students scored lower than the upper real limit of that interval, which is 94.5. The entries in the cumulative frequency distribution are obtained by a simple process of successive addition of the entries in the frequency column. Thus, the cumulative frequency corresponding to the upper real limit of the interval 104.5–109.5 is obtained by successive addition of $3 + 5 + 5 + 4 + 12 + 14 = 43$. Note that the top entry in the cumulative frequency column is always equal to N. If you fail to obtain this result, you know that you have made an error in cumulating frequencies and should check your work.

Table 3.4

Grouped frequency distribution and cumulative frequency distribution based upon data appearing in Table 3.3. $N = 110$

Class interval	f	Cumulative f	Cumulative %
150–154	2	110	100
145–149	2	108	98
140–144	3	106	96
135–139	5	103	94
130–134	7	98	89
125–129	9	91	83
120–124	9	82	75
115–119	13	73	66
110–114	17	60	55
105–109	14	43	39
100–104	12	29	26
95–99	4	17	15
90–94	5	13	12
85–89	5	8	7
80–84	3	3	3

The cumulative percentage distribution, also shown in Table 3.4, is obtained by dividing each entry in the cumulative frequency column by N and multiplying by 100. Note that the top entry must be 100% since all cases fall below the upper real limit of the highest interval.

3.3 GRAPHING TECHNIQUES

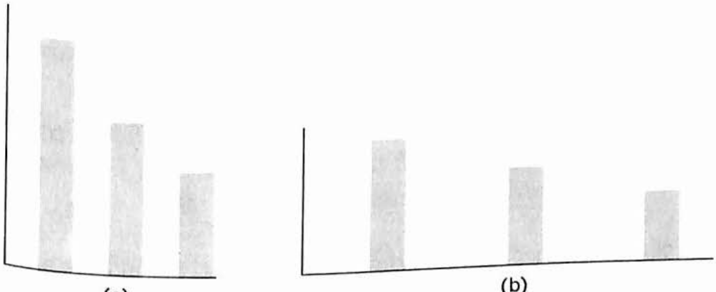
We have just examined some of the procedures involved in making sense out of a mass of unorganized data. As we pointed out, your work is usually just beginning when you have constructed frequency distributions of data. The next step, commonly, is to present the data in pictorial form so that the reader may readily apprehend the essential features of a frequency distribution and compare one with another if he desires. Such pictures, called graphs, should *not* be thought of as substitutes for statistical treatment of data but rather as *visual aids* for thinking about and discussing statistical problems.

3.4 MISUSE OF GRAPHING TECHNIQUES

As you are well aware, graphs are often employed in the practical world of commerce to mislead the reader. For example, by the astute manipulation of the vertical (ordinate or Y-axis) and horizontal (abscissa or X-axis) axes of a graph, it is possible to convey almost any impression that is desired. Figure 3.1 illustrates this misapplication of graphing techniques. In it are shown two bar graphs of the same data in which the ordinate and the abscissa are successively elongated to produce two distinctly different impressions.

It will be noted that graph (a) tends to exaggerate the difference in frequency counts among the four classes, whereas graph (b) tends to minimize these differences.

The differences might be further exaggerated by use of a device called the "Gee whiz!" chart by Daryl Huff in his excellent book, *How to Lie with Statistics.*



(a) (b)

Fig. 3.1 Bar graphs representing the same data but producing different impressions by varying the relative lengths of the ordinate and the abscissa.

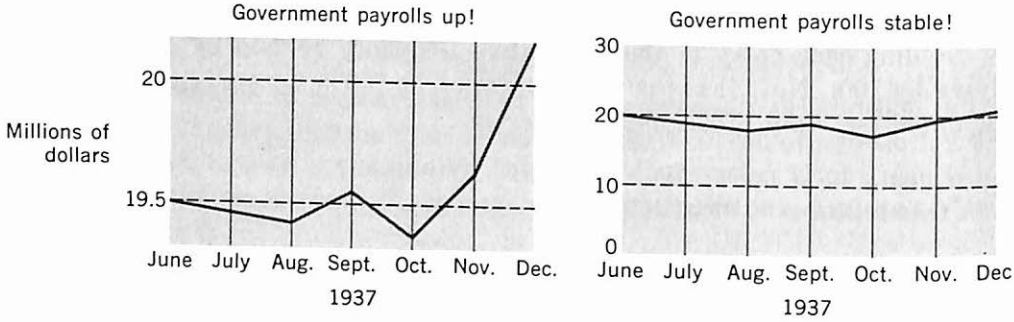


Fig. 3.2 The use of the "Gee Whiz" chart to exaggerate differences along the ordinate (Reprinted from D. Huff, *How to Lie with Statistics.* New York: W. W. Norton & Co., Inc., 1954, with permission).

This procedure consists of eliminating the zero frequency from the vertical axis and beginning with a frequency count greater than zero. Figure 3.2 is borrowed from Mr. Huff's book and illustrates quite dramatically the way in which graphs may be employed for purposes of deception.

It is obvious that the use of such devices are inimical to the aims of the statistician which are to present data with such clarity that misinterpretations are reduced to a minimum. We may overcome the second source of error illustrated above by making the initial entry on the Y-axis a zero frequency. The first problem, however, which is the selection of scale units to represent the horizontal and vertical axes, remains. Clearly, the choice of these units is an arbitrary affair, and anyone who decides to make the ordinate twice the length of the abscissa is just as correct as one who decides upon the opposite representation. It is clear that in order to avoid graphic anarchy, it is necessary to adopt a convention.

3.4.1 Three-Quarter High Rule

For graphic representations of frequency distributions, most statisticians have agreed upon a convention known as the "three-quarter high rule" which is expressed as follows.

> *In plotting the frequencies, the vertical axis should be laid out so that the height of the maximum point (representing the score with the highest associated frequency) is approximately equal to three quarters the length of the horizontal axis.*

The advantage of this convention is that it prevents subjective factors and, possibly, personal biases from influencing decisions concerning the relative proportion of the abscissa and the ordinate in graphic representations. The use

of the three-quarter rule is illustrated in the forthcoming section dealing with bar graphs. However, this rule has also been applied to all the graphs appearing in the remainder of the chapter.

3.5 NOMINALLY SCALED VARIABLES

The bar graph, illustrated in Fig. 3.3, is the graphic device employed to represent data which are either nominally or ordinally scaled. A vertical bar is drawn for each category in which the height of the bar represents the number of members of that class. If we arbitrarily set the width of each bar at one unit, the *area* of each bar may be used to represent the frequency for that category. Thus the total area of all the bars is equal to N.

In preparing frequency distributions of nominally scaled variables, you must keep two things in mind: (1) No order is assumed to underlie nominally scaled variables. Thus, the various categories can be represented along the abscissa in any order you choose. The authors prefer to arrange the categories alphabeti-

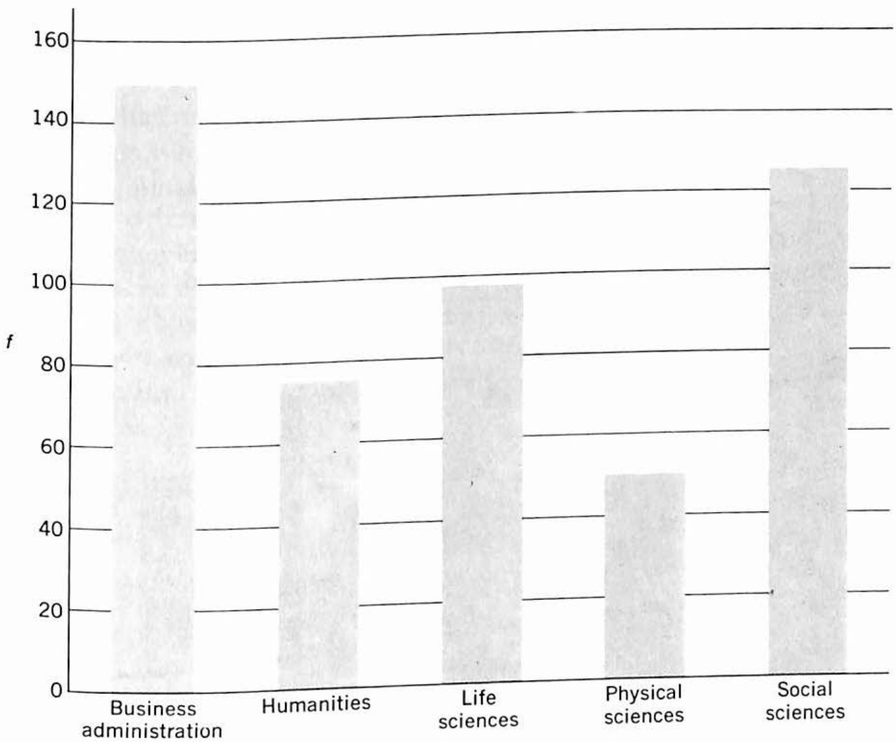


Fig. 3.3 Number of students enrolled in introductory economics courses who are majoring in the various academic fields (hypothetical data).

cally in keeping with their desire to eliminate any possibility of personal factors entering into the decision. (2) The bars should be separated rather than contiguous so that any implication of continuity among the categories is avoided.

The use of the three-quarter rule is perfectly straightforward with graphs of nominally scaled variables. Parenthetically, in implementing the three-quarter rule, we recommend that you acquaint yourself with the use of the metric ruler. You will soon find that the metric system of measurement is far superior to the English system when it comes to dealing with fractions and decimals.

The first decision which must be made concerning the length of the abscissa will be based on such factors as the amount of space available for the graphic representation. Once this decision is made, the remainder follows automatically.

Let us say that available space permits us to represent the horizontal axis with a line approximately 100 millimeters (mm) in length. The height of the vertical axis would then be $\frac{3}{4} \times 100$ or 75 mm. In Fig. 3.3, the maximum frequency is 150. Since 75 mm are to be shared by 150 frequencies, the number of millimeters representing each frequency is $\frac{75}{150}$ or 0.5 mm. In other words, each 5 mm along the Y-axis represents ten frequencies.

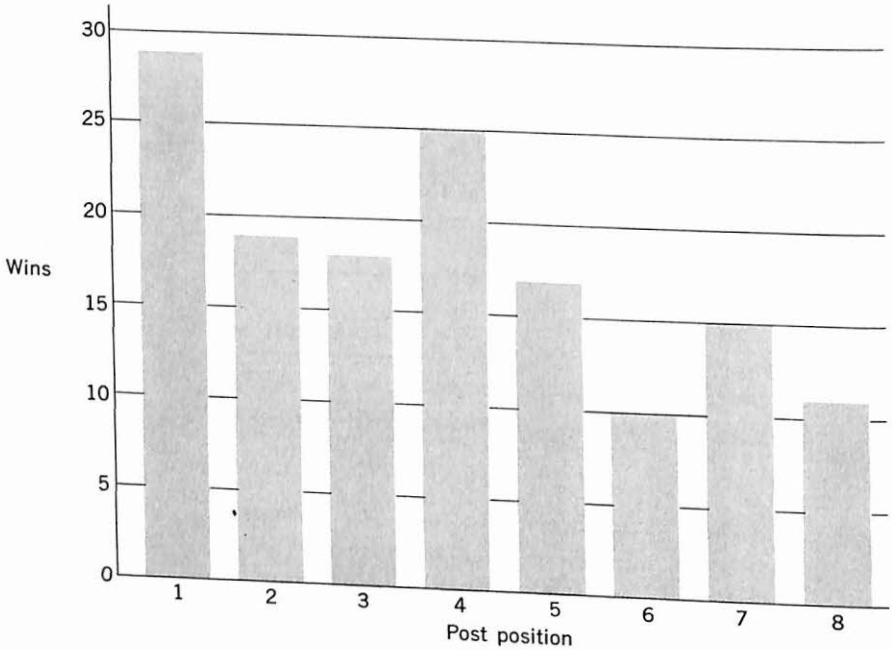


Fig. 3.4 Wins accrued on a circular track by horses from eight post positions. [From S. Siegel, *Non-Parametric Statistics*. New York: McGraw-Hill, 1956. Adapted with permission.]

Finally, the five categories represented along the abscissa will occupy 100 mm or 20 mm per class. It will be recalled that the vertical bars should be separated to avoid the implication of continuity. Consequently, the vertical bars should be somewhat less than 20 mm in width to permit this separation.

3.6 ORDINALLY SCALED VARIABLES

It will be recalled that the scale values of ordinal scales carry the implication of an ordering which is expressible in terms of the algebra of inequalities (greater than, less than). In terms of our preceding discussion, ordinally scaled variables should be treated in the same way as nominally scaled variables except that the categories should be placed in their naturally occurring order along the abscissa. Figure 3.4 illustrates the use of the bar graph with an ordinally scaled variable.

3.7 INTERVAL AND RATIO SCALED VARIABLES

3.7.1 Histogram

It will be recalled that interval and ratio scaled variables differ from ordinally scaled variables in one important way, i.e., equal differences in scale values are equal. This means that we may permit the vertical bars to touch one another in graphic representations of interval or ratio scaled frequency distributions. Such a graph is referred to as a histogram and replaces the bar graph employed with nominal and ordinal variables. Figure 3.5 illustrates the use of the histogram with a discretely distributed ratio scaled variable.

We previously noted (Section 3.5) that frequency may be represented either by the area of a bar or by its height. However, there are many graphic applications in which the height of the bar, or the ordinate, may give misleading information concerning frequency. Consider Fig. 3.6, which shows the data grouped into *unequal* class intervals and the resulting histogram.

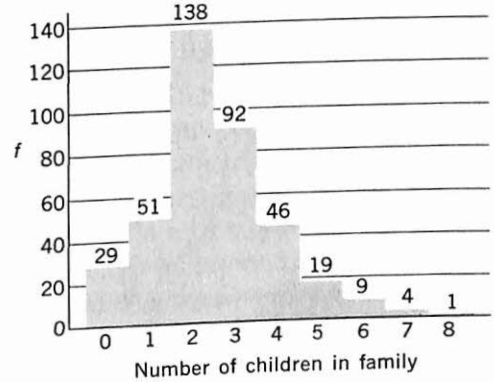
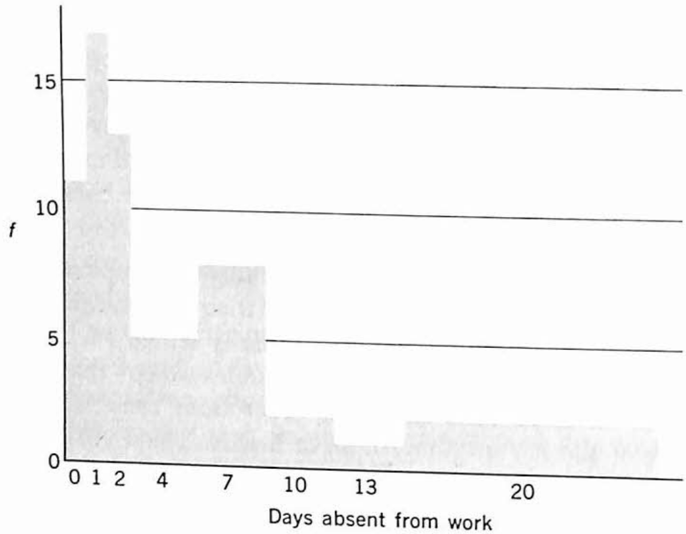


Fig. 3.5 Frequency distribution of the number of children per family among 389 families surveyed in a small suburban community (hypothetical data).



Number of days absent	Frequency
15–25	22
12–14	3
9–11	6
6–8	24
3–5	15
2	13
1	17
0	11

Fig. 3.6 Histogram employing unequal class intervals (hypothetical data).

If you think of frequency in terms of the height of the ordinate, you might erroneously conclude that the interval 15–25 includes only two cases. However, if we represent each score by one unit on the scale of frequency and an equal unit on the scale of scores, the total area for each score is equal to one. In the interval 15–25 there are 22 frequency units distributed over 11 score units; thus for this interval, the height of the ordinate will be $\frac{22}{11}$ or 2 score units. Similarily, in the interval 6–8 there are 24 frequency units distributed over 3 score units. Thus, the height of the ordinate must equal 8 units.

In general, it is advisable that we consider frequency in terms of area whenever we are dealing with variables in which an underlying continuity may be assumed.

3.7.2 Frequency Polygon

We can readily convert the histogram into another commonly employed form of graphic representation, the *frequency polygon*, by joining the midpoints of the bars with straight lines. However, it is not necessary to construct a histogram prior to the construction of a frequency polygon. All you need to do is place a dot where the tops of the bars would have been and join these dots. In practice, the authors prefer to reserve the use of the histogram for discrete distributions and the frequency polygon for distributions in which underlying continuity is explicit or may be assumed. When two or more frequency distributions are compared, the frequency polygon provides a clearer picture.

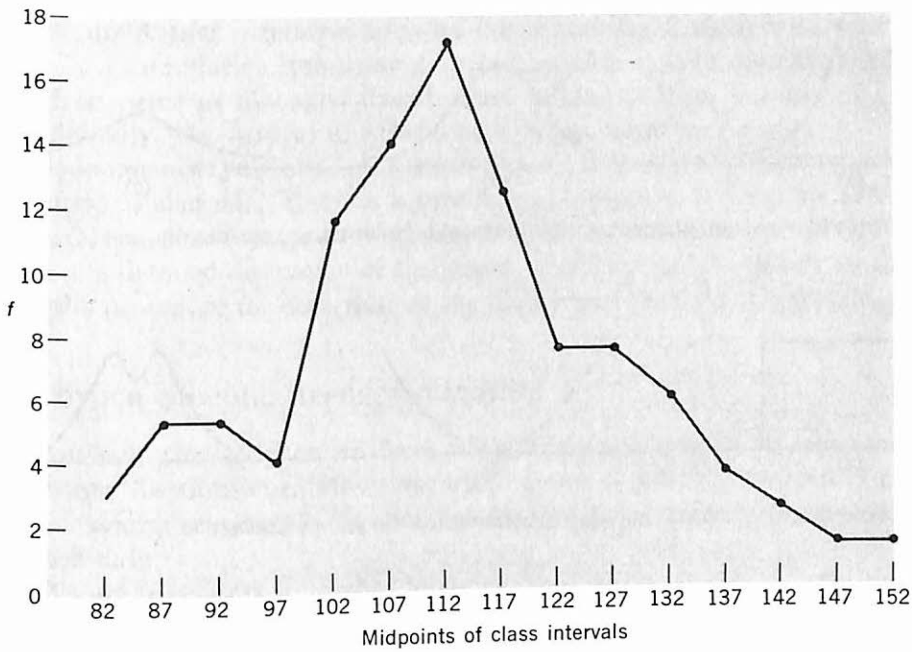


Fig. 3.7 Frequency polygon based on the data appearing in Table 3.4.

Figure 3.7 shows a frequency polygon based upon the grouped frequency distribution appearing in Table 3.3.

In frequency polygons, frequency *cannot* be represented by the height of the ordinate, but only by the area between the ordinates.

3.8 FORMS OF FREQUENCY CURVES

Frequency polygons may take on an unlimited number of different forms. However, many of the statistical procedures discussed in the text assume a particular form of distribution, namely, the "bell-shaped" normal curve.

In Fig. 3.8, several forms of bell-shaped distributions are shown. Curve (a), which is characterized by a piling up of scores in the center of the distribution, is referred to as a *leptokurtic* distribution. In curve (c), in which the opposite condition prevails, the distribution is referred to as *platykurtic*. And finally, curve (b) takes on the ideal form of the normal curve and is referred to as a *mesokurtic* distribution.

The normal curve is referred to as a symmetrical distribution, since, if it is folded in half, the two sides will coincide. Not all symmetrical curves are bell-shaped, however. A number of different symmetrical curves are shown in Fig. 3.9.

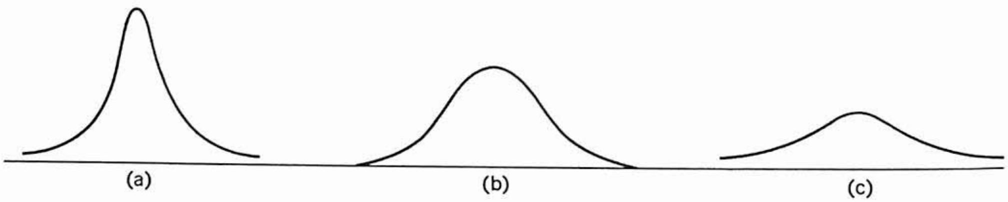


Fig. 3.8 Three forms of bell-shaped distributions: (a) leptokurtic, (b) mesokurtic, and (c) platykurtic.

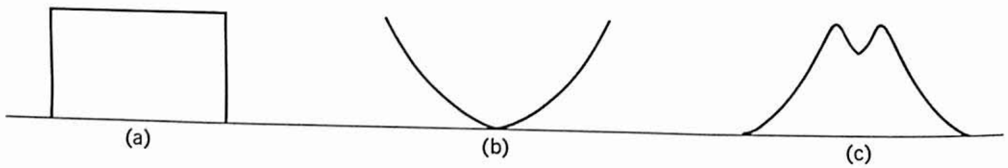


Fig. 3.9 Illustrations of several nonnormal symmetrical frequency curves.

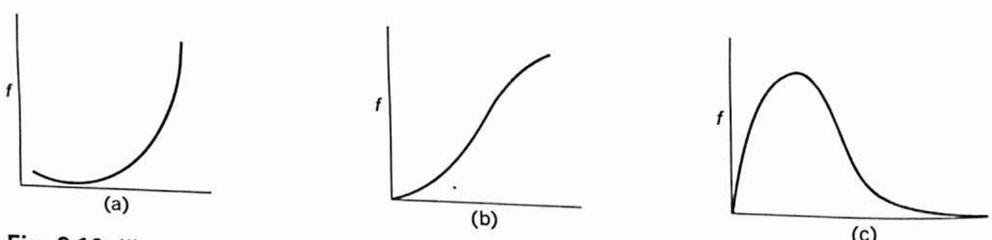


Fig. 3.10 Illustration of skewed frequency curves.

Certain distributions have been given names, e.g., that in Fig. 3.9(a) is called a *rectangular* distribution and that in Fig. 3.9(b) a U-distribution. Incidentally, a bimodal distribution such as appears in Fig. 3.9(c) is found when the frequency distributions of two different populations are represented in a single graph. For example, a frequency distribution of male and female adults of the same age would probably yield a curve similar to Fig. 3.9(c) on a strength of grip task.

When a distribution is not symmetrical, it is said to be *skewed*. If we say that a distribution is positively skewed, we mean that the distribution tails off at the high end of the horizontal axis and there are relatively fewer frequencies at this end. If, on the other hand, we say that the distribution is negatively skewed, we mean that there are relatively fewer scores associated with the left-hand, or low, side of the horizontal axis. Figure 3.10 presents several forms of skewed distributions.

Figure 3.10(a) is referred to as a J-curve and Fig. 3.10(b) is referred to as an ogive. A cumulative frequency distribution of normally distributed data will yield an ogive or S-shaped distribution. Figure 3.10(c) is positively skewed. Incidentally, Fig. 3.10(a) illustrates an extreme negative skew.

It is not always possible to determine by inspection whether or not a distribution is skewed. There is a precise mathematical method for determining both direction and magnitude of skew. It is beyond the scope of this book to go into a detailed discussion of this topic. In Chapter 4, however, we shall outline the procedure for determining the direction, if not the magnitude, of skew.

3.9 OTHER GRAPHIC REPRESENTATIONS

Throughout this chapter we have been discussing graphic representations of frequency distributions. However, other types of data are frequently collected by behavioral scientists. We shall briefly discuss a few graphic representations of such data.

We are frequently interested in comparing various groups or conditions with respect to a given characteristic. These groups or conditions constitute the independent, or experimental, variable, whereas the characteristic we are measuring is referred to as the criterion or dependent variable. Figure 3.11 compares

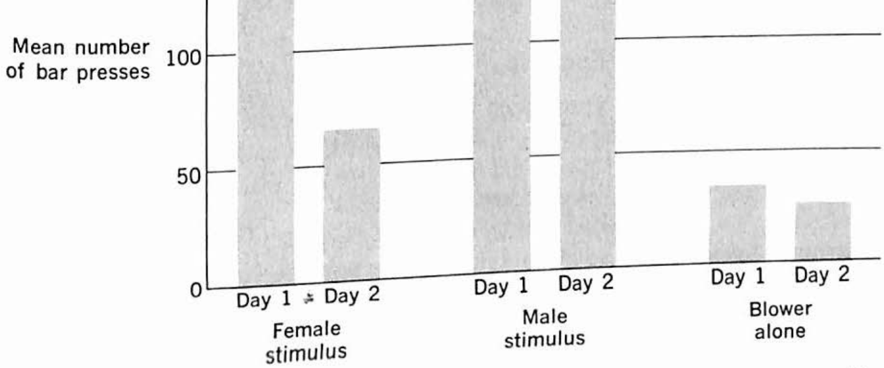


Fig. 3.11 Mean number of bar presses among three groups of white rats (Runyon and Kosacoff, 1965).

the mean number of bar pressing responses among three groups of animals receiving three types of olfactory reinforcement, e.g., odors of female rats, odors of male rats, and the control of blower alone (Runyon and Kosacoff, 1965). It will be noted that the experimental variable is represented along the X-axis and the criterion or dependent variable on the Y-axis.

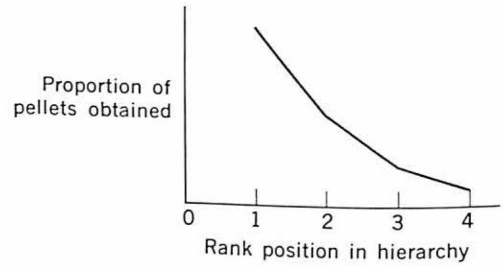


Fig. 3.12 Proportion of food pellets obtained by animals at each position in the NFT hierarchy (for nine animals at each position) (Runyon and Turner, 1964).

In Fig. 3.12 we see a line graph showing the relationship between two variables in a social structure study (Runyon and Turner, 1964). In this experiment, four animals were placed in a highly competitive quasi-natural environment, and various behavioral dimensions were studied. To determine whether one measure, nosing of the food tray (NFT), was a valid measure of food control, the animals were rank ordered from highest to lowest in terms of the number of NFT's observed. The number of food pellets obtained by the animals at each rank position was then calculated. Figure 3.12 reveals a clear-cut decrease in the number of food pellets obtained by the animals that were successively lower on the NFT hierarchy.

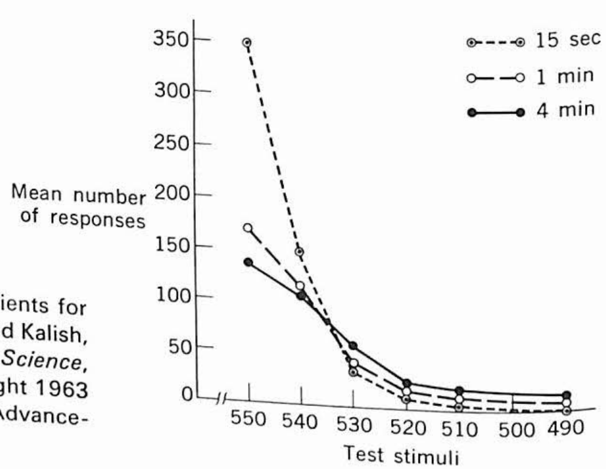


Fig. 3.13 Mean generalization gradients for three reinforcement groups (Haber and Kalish, 1963). (Reprinted by permission from *Science*, **142**, 18 Oct. 1963, p. 412. Copyright 1963 by the American Association for the Advancement of Science.)

Figure 3.13 presents another line graph, this time representing data taken from three different reinforcement groups. In this study (Haber and Kalish, 1963) groups of animals were trained to respond to a 550-mμ stimulus under three different levels of reinforcement, and were then tested for generalization of the response by presenting stimuli along the continuum. The differential effects of various schedules of reinforcement were most evident in the differences among the gradients.

CHAPTER SUMMARY

This chapter was concerned with the techniques employed in "making sense" out of a mass of data. We demonstrated the construction of frequency distributions of scores and presented various graphing techniques. When the scores are widely spread out, many have a frequency of zero, and when there is no clear indication of central tendency, it is customary to group scores into class intervals. The resulting distribution is referred to as a grouped frequency distribution.

The basis for arriving at a decision concerning the grouping units to employ and the procedures for constructing a grouped frequency distribution were discussed and demonstrated. It was seen that the true limits of a class interval are obtained in the same way as the true limits of a score. The procedures for converting a frequency distribution into a cumulative frequency distribution and a cumulative percentage distribution were demonstrated.

We also reviewed the various graphing techniques employed in the behavioral sciences. The basic purpose of graphical representation is to provide visual aids for thinking about and discussing statistical problems. The primary objective is to present data in a clear, unambiguous fashion so that the reader may apprehend at a glance the relationships which we want to portray.

We discussed the following:

1. Devices employed by unscrupulous individuals to mislead the unsophisticated reader.
2. The use of the bar graph with nominally and ordinally scaled variables.
3. The use of the histogram and the frequency polygon with continuous and discontinuous ratio or interval scaled variables.
4. The use of the "three-quarter high" rule to standardize the representation of the abscissa and ordinate in graphic techniques.
5. Various forms of normally distributed data, nonnormal symmetrical distributions, and asymmetrical or skewed distributions.
6. The use of the pie (circle) chart to represent distribution ratios.
7. Finally, we discussed and demonstrated several graphic representations of data, other than frequency distributions, commonly employed in the behavioral sciences.

Terms to Remember:

Random
Frequency distribution
Grouped frequency distribution
Class intervals
Mutually exclusive
Abscissa (X-axis)
Ordinate (Y-axis)
Bar graph
"Gee Whiz" chart
"Three-quarter high rule"
Histogram
Frequency polygon
Normal curve
Platykurtic distribution
Leptokurtic distribution
True limits of an interval
Apparent limits of an interval
Cumulative frequency distribution
Cumulative percentage distribution
Mesokurtic distribution
Rectangular distribution
U-distribution
Ogive
Skew
J-curve
Positively skewed distribution
Negatively skewed distribution

EXERCISES

1. Give the true limits, the midpoints, and the width of interval for each of the following class intervals.
   a) 8–12 b) 6–7 c) 0–2 d) 5–14
   e) (—2)–(—8) f) 2.5–3.5 g) 1.50–1.75 h) (—3)–(+3)
2. For each of the following sets of measurements, state (a) the best width of class interval (i), (b) the apparent limits of the lowest interval, (c) the true limits of that interval, (d) the midpoint of that interval.
   i) 0 to 106 ii) 29 to 41 iii) 18 to 48
   iv) —30 to +30 v) 0.30 to 0.47 vi) 0.206 to 0.293
3. Given the following list of scores in a statistics examination, use $i = 5$ for the class intervals and (a) set up a frequency distribution; (b) list the true limits and the midpoint of each interval; (c) prepare a cumulative frequency distribution and (d) prepare a cumulative percentage distribution.

   Scores on a Statistics Examination

63	88	79	92	86	87	83	78	41	67
68	76	46	81	92	77	84	76	70	66
77	75	98	81	82	81	87	78	70	60
94	79	52	82	77	81	77	70	74	61

4. Using the data in Problem 3, set up frequency distributions with the following: (a) $i = 1$ (ungrouped frequency distribution), (b) $i = 3$, (c) $i = 10$, (d) $i = 20$. Discuss the advantages and the disadvantages of employing these widths.

5. Given the following list of numbers: (a) Construct a grouped frequency distribution. (b) List the true limits and the midpoint of each interval. Indicate the width employed. (c) Compare the results with those of Problems 3 and 4.

6.3	8.8	7.9	9.2	8.6	8.7	8.3	7.8	4.1	6.7
6.8	7.6	4.6	8.1	9.2	7.7	8.4	8.6	7.0	6.6
7.7	7.5	9.8	8.1	8.2	8.1	8.7	7.8	7.0	6.0
9.4	7.9	5.2	8.2	7.7	8.1	7.7	7.0	7.4	6.1

6. Several entries in a frequency distribution showing the yield of corn per acre of land are 15–21, 8–14, 1–7. (a) What is the width of the interval? (b) What are the lower and upper real limits of each interval? (c) What are the midpoints of each interval?

7. Listed in Table 3.5 are the high prices per share of 250 stocks sold on the New York Stock Exchange on January 27, 1968, as shown in The New York Times, January 27, 1968.
   a) Round each price to the nearest dollar and group the results into a frequency distribution.
   b) Try several grouping intervals and note changes on the form of the resulting frequency distributions.
   c) Identify the apparent and real limits of the lowest class interval for each of the resulting frequency distributions.

8. Construct a grouped frequency distribution, using 5–9 as the lowest class interval, for the following list of numbers. List the width, midpoint, and real limits of the highest class interval.

67	63	64	57	56	55	53	53	54	54
45	45	46	47	37	23	34	44	27	44
45	34	34	15	23	43	16	44	36	36
35	37	24	24	14	43	37	27	36	26
25	36	26	5	44	13	33	33	17	33

9. Do Problem 8 again, using 3–7 as the lowest class interval. Compare the resulting frequency distribution with those of Problems 8, 10, and 11.

10. Repeat Problem 8, using $i = 2$. Compare the results with those of Problems 8, 9, and 11.

11. Repeat Problem 8, using $i = 10$. Compare the results with those of Problems 8, 9, and 10.

12. Give an example of each of the following distributions:
    a) normal distribution
    b) U-shaped distribution
    c) positively skewed distribution
    d) negatively skewed distribution
    e) rectangular distribution
    f) bimodal distribution

Table 3.5
High prices per share of 250 stocks

55	$90\frac{7}{8}$	$22\frac{1}{2}$	$40\frac{1}{2}$	81	47	$48\frac{3}{8}$	9	$37\frac{1}{2}$	$26\frac{3}{8}$
$16\frac{1}{2}$	$21\frac{5}{8}$	$30\frac{1}{2}$	$68\frac{3}{4}$	68	$29\frac{5}{8}$	$31\frac{3}{4}$	$44\frac{1}{8}$	$20\frac{3}{4}$	$29\frac{1}{2}$
$73\frac{5}{8}$	$42\frac{3}{4}$	$26\frac{3}{8}$	22	$58\frac{1}{8}$	$55\frac{3}{4}$	$18\frac{1}{2}$	$31\frac{1}{2}$	$43\frac{1}{2}$	$30\frac{5}{8}$
$39\frac{3}{4}$	$21\frac{5}{8}$	$37\frac{7}{8}$	$25\frac{7}{8}$	$185\frac{1}{4}$	$47\frac{1}{8}$	$34\frac{3}{4}$	$34\frac{1}{2}$	$32\frac{1}{2}$	42
$10\frac{3}{8}$	$69\frac{5}{8}$	$64\frac{1}{4}$	$29\frac{1}{2}$	230	18	38	$62\frac{1}{8}$	$21\frac{3}{4}$	$106\frac{3}{4}$
$76\frac{1}{2}$	35	$44\frac{3}{4}$	$32\frac{7}{8}$	$97\frac{1}{4}$	26	34	$36\frac{1}{2}$	$24\frac{1}{8}$	$17\frac{1}{2}$
27	$53\frac{7}{8}$	$33\frac{1}{8}$	$34\frac{7}{8}$	76	$14\frac{5}{8}$	26	$72\frac{1}{8}$	$30\frac{3}{4}$	$21\frac{1}{2}$
30	$52\frac{3}{4}$	$90\frac{3}{4}$	$31\frac{5}{8}$	$51\frac{1}{2}$	$32\frac{1}{2}$	$33\frac{1}{4}$	$23\frac{3}{4}$	$23\frac{5}{8}$	33
$8\frac{5}{8}$	$36\frac{7}{8}$	$20\frac{5}{8}$	$28\frac{1}{4}$	$42\frac{5}{8}$	25	$104\frac{1}{2}$	$57\frac{3}{8}$	$53\frac{3}{4}$	63
$34\frac{1}{2}$	78	$12\frac{3}{4}$	$17\frac{7}{8}$	$43\frac{1}{4}$	$10\frac{3}{4}$	$46\frac{1}{2}$	$35\frac{1}{2}$	67	$39\frac{7}{8}$
$19\frac{7}{8}$	$89\frac{1}{8}$	$27\frac{5}{8}$	$53\frac{3}{4}$	$65\frac{7}{8}$	$10\frac{3}{4}$	$38\frac{1}{2}$	$25\frac{5}{8}$	$78\frac{1}{2}$	16
$12\frac{1}{8}$	66	27	$14\frac{7}{8}$	$14\frac{3}{4}$	40	$42\frac{3}{8}$	32	$120\frac{3}{8}$	$26\frac{1}{4}$
$11\frac{7}{8}$	$14\frac{1}{2}$	$55\frac{1}{2}$	$54\frac{1}{2}$	$28\frac{1}{8}$	$26\frac{3}{4}$	$57\frac{1}{2}$	$111\frac{5}{8}$	$131\frac{1}{2}$	22

$33\frac{1}{2}$	86	$40\frac{5}{8}$	$16\frac{3}{4}$	$15\frac{5}{8}$	$41\frac{5}{8}$	$78\frac{1}{4}$	$10\frac{7}{8}$	$14\frac{7}{8}$	$32\frac{1}{8}$
$36\frac{1}{2}$	$42\frac{1}{2}$	$33\frac{1}{8}$	$114\frac{3}{4}$	$55\frac{1}{8}$	32	$31\frac{5}{8}$	$31\frac{1}{2}$	13	$50\frac{3}{8}$
$42\frac{1}{4}$	56	$14\frac{1}{4}$	$24\frac{1}{8}$	$29\frac{1}{8}$	$64\frac{5}{8}$	$17\frac{1}{4}$	69	$12\frac{1}{2}$	$29\frac{5}{8}$
$33\frac{1}{4}$	$6\frac{1}{2}$	$14\frac{1}{2}$	$34\frac{3}{8}$	$27\frac{1}{4}$	$31\frac{1}{8}$	$80\frac{1}{2}$	$27\frac{7}{8}$	34	41
$21\frac{7}{8}$	$28\frac{1}{2}$	$107\frac{1}{4}$	24	77	70	$60\frac{5}{8}$	91	$54\frac{1}{2}$	$37\frac{3}{4}$
$28\frac{3}{4}$	$16\frac{7}{8}$	101	$21\frac{1}{4}$	86	36	$37\frac{1}{4}$	$24\frac{5}{8}$	$54\frac{5}{8}$	$57\frac{5}{8}$
$11\frac{1}{8}$	$54\frac{1}{4}$	51	$67\frac{3}{4}$	$61\frac{5}{8}$	13	34	48	$15\frac{1}{4}$	$40\frac{1}{8}$
$34\frac{3}{4}$	93	40	$35\frac{3}{8}$	$22\frac{1}{4}$	$50\frac{3}{8}$	8	$32\frac{1}{2}$	$57\frac{1}{8}$	$43\frac{1}{2}$
$41\frac{1}{2}$	$23\frac{5}{8}$	$26\frac{1}{4}$	96	$49\frac{1}{2}$	$32\frac{5}{8}$	$9\frac{3}{4}$	108	85	$29\frac{3}{8}$
$23\frac{3}{4}$	$43\frac{1}{4}$	$64\frac{1}{2}$	$72\frac{5}{8}$	$43\frac{1}{2}$	$68\frac{7}{8}$	43	$163\frac{1}{4}$	$61\frac{7}{8}$	114
$16\frac{3}{8}$	$51\frac{1}{2}$	$51\frac{1}{4}$	24	$9\frac{3}{4}$	$92\frac{7}{8}$	$33\frac{5}{8}$	132	53	103
$56\frac{1}{4}$	$20\frac{3}{4}$	$45\frac{1}{8}$	84	29	$43\frac{1}{4}$	$29\frac{3}{4}$	107	85	33

13. Given the following frequency distribution of the weights of 96 female students, draw a histogram.

Class interval	f	Class interval	f
160–164	1	130–134	17
155–159	3	125–129	11
150–154	10	120–124	8
145–149	6	115–119	3
140–144	14	110–114	1
135–139	22		

14. Draw the appropriate graphic representations for each of the four examples in Chapter 2, Problem 9, using your own hypothetical data.

15. Take a pair of dice, toss 100 times and record the sum (on the face of the two dice) for each toss. Prepare a bar graph, showing the number of times each sum occurs.

16. Given the following monthly sales by five salesmen in a large appliance store:

Salesmen	Sales, \$	Salesmen	Sales \$
Mr. Art	22,500	Mr. Warren	22,100
Mr. Harold	17,900	Mr. Cye	20,700
Mr. Marty	21,400		

Draw graphs to perpetrate the lies stated below.

a) The sales manager wants to impress the owner of the store that all members of his sales force are functioning at a uniformly high level.
b) The sales manager wants to spur Mr. Harold to greater efforts.
c) The store owner wants to spur the sales manager to greater efforts.

17. Draw a graph of the data in Problem 16 which represents the true state of affairs.

18. Below are the scores of two groups of fourth-grade students on a test of reading ability.

Class interval	Group A f	Group B f	Class interval	Group A f	Group B f
50–52	5	2	29–31	9	22
47–49	12	3	26–28	6	11
44–46	18	5	23–25	4	9
41–43	19	8	20–22	3	6
38–40	26	12	17–19	1	4
35–37	19	24	14–16	2	2
32–34	13	35			
				137	143

a) Construct a frequency polygon for each of these groups on the same axis.
b) Describe and compare each distribution.

19. Describe the types of distributions you would expect if you were to graph each of the following.
    a) Annual incomes of American families.
    b) The heights of adult American males.
    c) The heights of adult American females.
    d) The heights of American males and females combined in one graph.
20. Given the following frequency distribution of the results of 73 students on a midterm exam, draw a frequency polygon.

Class interval	f	Class interval	f
95–99	2	65–69	11
90–94	2	60–64	6
85–89	5	55–59	3
80–84	9	50–54	0
75–79	16	45–49	1
70–74	18		

21. For the data in Problem 20, draw a cumulative frequency polygon.
22. Obtain the annual budget estimate of the President of the United States for this year. Show the allocation of funds by category, using both a bar graph and a pie chart.
23. Carefully note the forms of graphical representation employed in the financial section of your Sunday newspaper. Collect examples of "good" graphics, i.e., clear to interpret and free of misleading elements. Collect examples of "poor" graphics. What makes them poor? Begin to list characteristics which determine whether graphs are effective or ineffective. Note also graphs employed in advertising. How do they compare in quality to the graphs illustrating financial statistics. Are they as complete? Are the ordinate and abscissa always labeled? Are any misleading? In what way?
24. Following are the number of quarts of milk sold at a supermarket on 52 consecutive Saturdays.

67	56	64	78	88	57	88	75
65	70	66	67	67	71	77	62
61	60	72	69	73	76	64	64
69	75	68	65	66	63	69	
65	62	71	72	61	64	73	
78	63	58	63	64	65	58	
75	62	71	92	74	65	61	

a) The manager of the dairy department decides to limit the number of quarts of milk on sale each Saturday to 70. Assuming that the sales figures will be the same the following year, what is the likelihood (i.e., the percentage of time) that his department will be caught short.
b) Group the sales figures into a frequency distribution with the lower class limits of 56–58.
c) Prepare a cumulative frequency distribution and cumulative percentage distribution based on the above frequency distribution.
d) Group the sales figures into a frequency distribution with the lower class limit of 55–59. Prepare a cumulative frequency distribution and cumulative percentage distribution based on this frequency distribution. Compare the resulting distributions with those obtained above.

25. Draw a frequency polygon for the frequency distribution obtained in Problem 8.
26. Draw a frequency polygon for the frequency distribution obtained in Problem 9. Compare this distribution with that of Problem 25.
27. Draw a cumulative frequency polygon for the frequency distribution obtained in Problem 8.
28. Draw a cumulative frequency polygon for the frequency distribution obtained in Problem 9. Compare this distribution with that of Problem 27.

Percentiles 4

4.1 INTRODUCTION

Let us suppose that a younger brother or sister came home from school and announced, "I received a score of 127 on my scholastic aptitude test." What would be your reaction? Commend him for obtaining such a fine score? Criticize him for not getting a higher one? Or reserve judgment until you learned more about the distribution of scores within your sibling's class or group? If you have passed the course up to this point, you have undoubtedly selected the last of these three alternatives.

It should be clear that a score by itself is meaningless. It takes on meaning only when it can be compared to some standard scale or base. Thus, if your younger sibling were to volunteer the information, "79% of the students scored lower than I," he would be providing some frame of reference for interpreting the score. Indeed, he would have been citing the percentile rank of his score. The *percentile rank* of a score, then, *represents the percent of cases in a comparison group which achieved scores lower than the one cited*. Thus to say that a score of 127 has a percentile rank of 79 is to indicate that 79% of the comparison group scored below 127. Incidentally, each score is considered to be a hypothetical point without dimension, so that it would be equally meaningful to say that 21% of the comparison group scored higher than 127.

4.2 CUMULATIVE PERCENTILES AND PERCENTILE RANK

4.2.1 Obtaining the Percentile Rank of Scores from a Cumulative Percentage Graph

In Chapter 3, we learned how to construct cumulative frequency and cumulative percentage distributions. If we were to graph a cumulative percentage distribution, we could read the percentile ranks directly from the graph.*

* Note that the reverse is also true, i.e., given a percentile rank, we could read the corresponding scores.

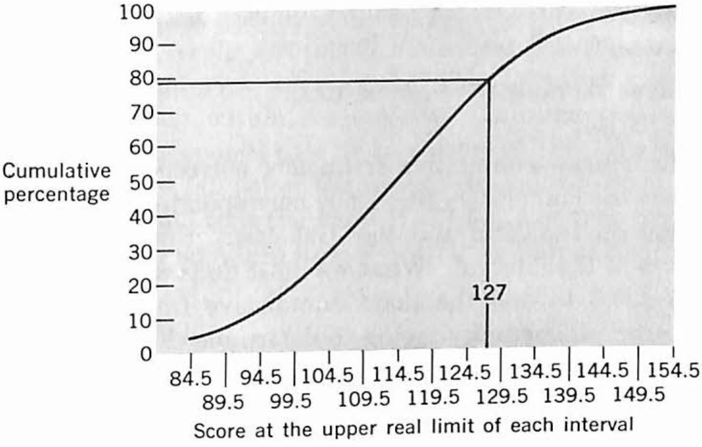


Fig. 4.1 Graphic representation of a cumulative percentage distribution.

Figure 4.1 presents a graphic form of the cumulative percentage distribution presented in Table 3.4.

To illustrate, let us imagine that we wanted to determine the percentile rank of a score of 127. We locate 127 along the abscissa and construct a perpendicular at that point so that it intercepts the curve. Reading directly across on the scale to the left, we see that the percentile rank is approximately 79. On the other hand, if we wanted to know the score at a given percentile, we could reverse the procedure. For example, what is the score at the 90th percentile? We locate the 90th percentile on the ordinate, we read directly to the right until it intercepts the curve, we construct a line perpendicular to the abscissa, and we read the value on the scale of scores. In the present example, it can be seen that the score at the 90th percentile is approximately 135.

4.2.2 Obtaining the Percentile Rank of Scores Directly

We are often called upon to determine the percentile rank of scores without the assistance of a cumulative percentage polygon or with greater precision than is possible with a graphical representation. To do this, it is usually necessary to interpolate within the cumulative frequency column to determine the precise cumulative frequency corresponding to a given score.

Using the grouped frequency distribution found in Table 4.1, let us determine directly the percentile rank of a score of 127 which we previously approximated by the use of the cumulative percentage polygon. The first thing we should note is that a score of 127 falls within the interval 125–129. The total cumulative frequency below that interval is 82. Since a percentile rank of a

score is defined symbolically as

$$\text{Percentile rank} = \frac{\text{cum } f}{N} \times 100, \tag{4.1}$$

it is necessary to find the precise cumulative frequency corresponding to a score of 127. It is clear that the cumulative frequency corresponding to a score of 127 lies somewhere between the 82nd and the 91st cases, the cumulative frequencies at both extremes of the interval. What we must do is to interpolate within the interval 124.5–129.5 to find the exact cumulative frequency of a score of 127. In doing this, we are actually trying to determine the proportion of distance that we must move into the interval in order to find the number of cases included up to a score of 127.

Table 4.1

Grouped frequency distribution and cumulative frequency distribution of scores in an educational test (hypothetical data)

Class interval	f	Cumulative f
150–154	2	110
145–149	2	108
140–144	3	106
135–139	5	103
130–134	7	98
125–129	9	91
120–124	9	82
115–119	13	73
110–114	17	60
105–109	14	43
100–104	12	29
95–99	4	17
90–94	5	13
85–89	5	8
80–84	3	3
	$N = 110$	

A score of 127 is 2.5 score units above the lower real limit of the interval (that is, $127 - 124.5 = 2.5$). Since there are 5 score units within the interval, a score of 127 is 2.5/5 of the distance through the interval. We now make a

very important assumption, i.e., *that the cases or frequencies within a particular interval are evenly distributed throughout that interval.* Since there are 9 cases within the interval, we may now calculate that a score of 127 is $2.5/5 \times 9$, or the 4.5th case within the interval. In other words, the frequency 4.5 in the interval corresponds exactly to a score of 127. We have already seen, however, that 82 cases fall below the lower real limit of the interval. Adding the two together, we find that the score of 127 has a cumulative frequency of exactly 86.5. Substituting 86.5 into formula (4.1), we obtain the following:

$$\text{Percentile rank of } 127 = \frac{86.5}{110} \times 100 = 78.64.$$

You will note that this answer, when rounded to the nearest percentile, agrees with the approximation obtained by the use of the graphical representation of a cumulative percentage distribution (Fig. 4.1).

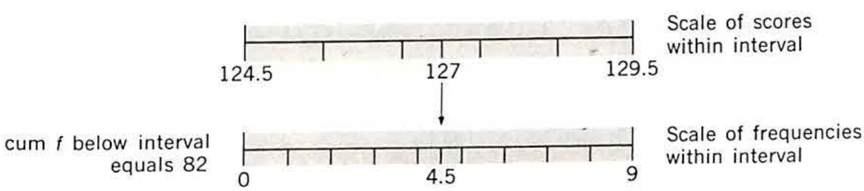


Fig. 4.2 Graphic representation of the procedures involved in finding the cumulative frequency corresponding to a given score.

Figure 4.2 summarizes graphically the procedures involved in finding the cumulative frequency of a given score. You will note that the interval 124.5–129.5 is divided into 5 equal units corresponding to the scores within that interval, whereas the frequency scale is divided into 9 equal units corresponding to the 9 frequencies within that interval. What we are accomplishing, in effect, in finding the frequency corresponding to a score, is a *linear transformation* from a scale of scores to a scale of frequencies, which is analogous to converting Fahrenheit readings to values on a centigrade scale and conversely.

Formula (4.2) presents a generalized formula for calculating the percentile rank of a given score.

$$\text{Percentile rank} = \frac{\text{cum } f_{ll} + \left(\dfrac{X - X_{ll}}{i}\right)(f_i)}{N} \times 100, \tag{4.2}$$

where $\text{cum } f_{ll}$ = cumulative frequency at the lower real limit of the interval containing X,
X = given score,
X_{ll} = score at lower real limit of interval containing X,
i = width of interval,
f_i = number of cases within the interval containing X.

4.2.3 Finding the Score Corresponding to a Given Percentile Value

Let us imagine that your younger brother, instead of apprising you of his score, reported, instead, his percentile rank. He tells you his score was at the 96th percentile. What was his score?

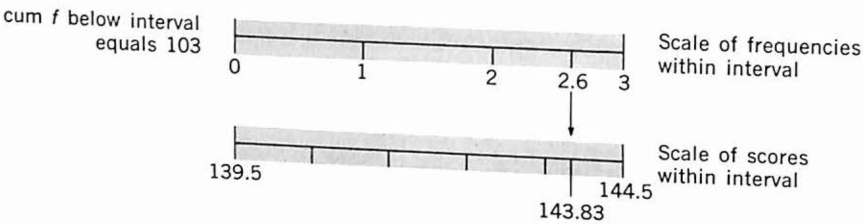


Fig. 4.3 Graphic representation of the procedures involved in finding the score corresponding to a given frequency within the interval.

To obtain the answer, we must interpolate in the reverse direction, from the cumulative frequency scale to the scale of scores. The first thing we must learn is the cumulative frequency corresponding to the 96th percentile. It follows algebraically from formula (4.1) that

$$\text{cum } f = \frac{\text{percentile rank} \times N}{100}. \tag{4.3}$$

Since we are interested in a score at the 96th percentile and our N is 110, the cumulative frequency of a score at the 96th percentile is

$$\text{cum } f = \frac{96 \times 110}{100} = 105.6.$$

Referring to Table 4.1, we see that the frequency 105.6 is in the interval with the real limits of 139.5–144.5. Indeed, it is 2.6 frequencies into the interval

since the cum f at the lower real limit of the interval is 103, which is 2.6 less than 105.6. There are three cases in all within the interval. Thus the frequency 105.6 is 2.6/3 of the way through an interval with a lower real limit of 139.5 and an upper real limit of 144.5. In other words, it is 2.6/3 of the way through 5 score units. Expressed in terms of score units, then, it is $(2.6/3) \times 5$ or 4.33 score units above the lower real limit of the interval. By adding 4.33 to 139.5, we obtain the score at the 96th percentile, which is 143.83.

Figure 4.3 represents, graphically, the procedures involved in the linear transformation from units of the frequency scale to units of the scale of scores. For the students desiring a generalized method for determining scores corresponding to a given percentile, formula (4.4) should be helpful.

$$\text{Score at a given percentile} = X_{ll} + \frac{i\,(\text{cum } f - \text{cum } f_{ll})}{f_i}, \tag{4.4}$$

where X_{ll} = score at lower real limit of the interval containing cum f,
i = width of the interval,
cum f = cumulative frequency of the score,
cum f_{ll} = cumulative frequency at the lower real limit of the interval containing cum f,
f_i = number of cases within the interval containing cum f.

To illustrate the use of the formula, let us employ an example with which we are already familiar. What score is at the 78.64th percentile? First, by employing formula (4.3) we obtain

$$\text{cum } f = \frac{78.64 \times 110}{100} = 86.50.$$

The score at the lower real limit of the interval containing the frequency 86.50 is 124.5; i is 5; cum f to the lower real limit of the interval is 82, and the number of cases within the interval is 9. Substituting the above values into formula (4.4), we obtain:

$$\begin{aligned}\text{score at 78.64 percentile} &= 124.5 + \frac{5(86.50 - 82)}{9}\\ &= 124.5 + 2.50 = 127.00.\end{aligned}$$

Note that this is the score from which we previously obtained the percentile rank 78.64 and that this formula illustrates, incidentally, a good procedure for checking the accuracy of our calculations. In other words, whenever you find the percentile rank of a score, you may take that answer and determine the score corresponding to that percentile value. You should obtain the original score. Similarly, whenever you obtain a score corresponding to a given per-

centile rank you may take that answer and determine the percentile rank of that score. You should always come back to the original percentile rank. Failure to do so indicates that you have made an error. It is preferable to repeat the solution without reference to your prior answer rather than to attempt to find the mistake in your prior solution. Such errors are frequently of the "proof-reader" type which defy detection, are time consuming to locate, and are highly frustrating.

4.3 PERCENTILE RANK AND REFERENCE GROUP

Just as a score is meaningless in the abstract, so also is a percentile rank. A percentile rank must always be expressed in relation to some reference group. Thus if a friend claims that he obtained a percentile rank of 93 in a test of mathematical aptitude, you might not be terribly impressed if the reference group were made up of individuals who completed only the eighth grade. On the other hand, if the reference group consisted of individuals holding a doctorate in mathematics, your attitude would unquestionably be quite different.

Many standardized tests employed in psychology, education, and industry publish separate norms for various reference groups. Table 4.2 shows the raw-score equivalents for selected percentile points on a test widely employed for graduate-school admissions. You will note that a person obtaining a raw score of 50 on this test, would obtain a percentile rank of 25, 35, 40, and 50 when compared successively with reference groups in the biological sciences, medical science, agriculture, and social work.

CHAPTER SUMMARY

In this chapter we saw that a score, by itself, is meaningless unless it is compared to a standard base or scale. Scores are often converted into units of the percentile rank scale in order to provide a readily understandable basis for their interpretation and comparison.

We saw that:

1. Percentile ranks of scores and scores corresponding to a given percentile may be approximated from a cumulative percentage graph.
2. Direct computational methods permit a more precise location of the percentile rank of a score and the score corresponding to a given percentile. These methods were discussed and demonstrated in the text.
3. A percentile rank is meaningless in the abstract. It must always be expressed in relation to some reference group.

Terms to Remember:

Percentile rank
Percentile
Reference group

EXERCISES

1. Estimate the percentile rank of the following scores, employing Fig. 4.1.
   a) 104.5 b) 112 c) 134
2. Calculate the percentile rank of the scores in Problem 1, employing Table 4.1
3. Estimate the scores corresponding to the following percentiles, employing Fig. 4.1.
   a) 25 b) 50 c) 75
4. Calculate the scores corresponding to the percentiles in Problem 3, employing Table 4.1.
5. A younger sibling informs you he obtained a score of 130 on a standard vocabulary test. What additional information might you seek in order to interpret his score?
6. Refer back to Chapter 3, Problem 18.
   a) A student in group A obtained a score of 45 on the test of reading ability. What is his percentile rank in the group?
   b) What is the percentile rank of a student in group B who also obtained a score of 45?
   c) Combine both groups into an overall frequency distribution and obtain the percentile rank of a score of 45. What happens to the percentile rank of the student in group A? group B? Why?
7. If we were to place all the scores shown in Table 4.1 into a hat, what is the likelihood that, selecting at random, we would obtain:
   a) a score equal to or higher than 131?
   b) a score equal to or lower than 131?
   c) a score equal to or below 84?
   d) a score equal to or above 154?
   e) a score between 117 and 133?
   f) a score equal to or greater than 148, or equal to or less than 82?
8. The following questions are based on Table 4.2.
   a) John H. proudly proclaims that he obtained a "higher" percentile ranking than his friend, Howard. Investigation of the fact reveals that his score on the test was actually lower. Must it be concluded that John H. was lying, or is some other explanation possible?
   b) Jack H. obtained a percentile rank of 65 on the social work scale. What was his score? What score would he have had to obtain to achieve the same percentile rank on the physical sciences scale?

Table 4.2 Raw-score equivalents of selected percentile points on the Miller analogies test for eight graduate and professional school groups. (Reproduced by permission. Copyright by The Psychological Corporation, New York, N.Y. All rights reserved.)

Percentile	Physical sciences	Agriculture	Medical science	Biological sciences
99	93	89	92	88
95	91	86	83	87
90	88	77	78	86
85	85	72	76	80
80	82	67	74	76
75	80	64	71	70
70	78	61	67	68
65	76	59	64	67
60	74	57	60	65
55	70	56	58	63
50	68	54	57	61
45	65	51	55	58
40	63	50	53	55
35	60	48	50	53
30	58	43	47	52
25	55	40	45	50
20	51	37	43	48
15	47	34	41	47
10	43	31	34	41
5	39	26	30	37
1	28	5	24	28
N	251	125	103	84
Mean	66.7	53.6	57.6	61.5
SD	16.6	17.3	16.2	15.6

c) World Law School employs the Miller Analogies Test as an element of the admissions procedure. No applicant obtaining a percentile rank below 75 is considered for admission regardless of his other qualifications. Thus the 75th percentile might be called a "cutoff" point. Which reference group is most likely involved in the decision? What score constitutes the cutoff point for this distribution?

d) Foster Medical School also employs the 75th percentile as a cutoff point. Paul F. obtained a raw score of 62. What are his chances of being considered for admission?

9. Employing the frequency distribution obtained in Problem 8, Chapter 3, calculate the percentile ranks of the following scores:
a) 9.5 b) 17.5 c) 24 d) 32 e) 34.5 f) 52.5

10. Repeat Problem 9, employing the frequency distribution obtained in Problem 9, Chapter 3. Compare the results with those of Problems 9, 11, and 12.

Social sciences	Social work	Languages and literature	Law school freshmen	Percentile
90	81	87	84	99
85	76	84	79	95
82	67	80	73	90
79	64	76	66	85
76	61	74	63	80
74	60	73	60	75
69	58	68	58	70
67	57	66	55	65
64	54	65	53	60
63	52	61	51	55
61	50	59	49	50
58	47	56	47	45
56	46	53	45	40
53	45	51	42	35
51	41	46	40	30
49	39	43	37	25
46	37	41	35	20
44	32	38	32	15
39	27	35	30	10
32	22	29	25	5
18	9	7	18	1
229	116	145	558	N
60.2	49.4	57.7	49.6	Mean
16.0	15.2	17.4	16.1	SD

11. Repeat Problem 9, employing the frequency distribution obtained in Problem 10, Chapter 3. Compare the results with those of Problems 9, 10, and 12.
12. Repeat Problem 9, employing the frequency distribution obtained in Problem 11, Chapter 3. Compare the results with those of Problems 9, 10, and 11.
13. Employing the frequency distribution obtained in Problem 8, Chapter 3, find the scores corresponding to the following percentiles:
    a) 10 b) 20 c) 50 d) 60 e) 75.
14. Repeat Problem 13, employing the frequency distribution obtained in Problem 9, Chapter 3. Compare the results with those of Problems 13, 15, and 16.
15. Repeat Problem 13, employing the frequency distribution obtained in Problem 10, Chapter 3. Compare the results with those of Problems 13, 14, and 16.
16. Repeat Problem 13, employing the frequency distribution obtained in Problem 11, Chapter 3. Compare the results with those of Problems 13, 14, and 15.

Measures of Central Tendency 5

5.1 INTRODUCTION

One of the greatest sources of confusion among lay people and, perhaps, a cause for their suspicions that statistics is more of an art than a science revolves about the ambiguity in the use of the term "average." Unions and management speak of average salaries, and frequently cite numerical values which are in sharp disagreement with each other; television programs and commercials are said to be prepared with the average viewer in mind; politicians are deeply concerned about the views of the average American voter; the average family size is frequently given as a fractional value, a statistical abstraction which is ludicrous to some and an absurdity to others; the term "average" is commonly used as a synonym for the term "normal;" the TV weatherman tells us we had an average day or that rainfall for the month is above or below average. Indeed, the term "average" has so many popular connotations that many statisticians prefer to drop it from the technical vocabulary and refer, instead, to *measures of central tendency*. We shall define a measure of central tendency as an *index of central location employed in the description of frequency distributions*. Since the center of a distribution may be defined in different ways, there will be a number of different measures of central tendency. In this chapter, we shall concern ourselves with three of the most frequently employed measures of central tendency: the mean, the median, and the mode.

5.1.1 Why Describe Central Tendency?

Through the first five chapters of the book, we concerned ourselves primarily with organizing data into a meaningful and useful form. Beyond this, however, we want to describe our data in such ways that quantitative statements can be made. A frequency distribution represents an organization of data but it does not, in itself, permit us to make quantitative statements either describing the distribution or comparing two or more distributions.

There are two features of many frequency distributions which statisticians have noted and have developed quantitative methods for describing: (1) Frequently data cluster around a central value which is between the two extreme values of the variable under study. (2) The data may tend to be dispersed and distributed about the central value in a way which can be specified quantitatively. The first of these features is the topic of the present chapter, whereas the second (dispersion) is discussed in the forthcoming chapter.

Being able to locate a point of central tendency, particularly when coupled with a description of the dispersion of scores about that point, can be very useful to the behavioral scientist. For example, he may be able to reduce a mass of data to a simple quantitative value which may be understood and communicated to other scientists.

We have already stated that the behavioral scientist is frequently called upon to compare the measurements obtained from two or more groups of subjects for the purpose of drawing inferences about the effects of an independent variable. Measures of central tendency greatly simplify the task of drawing conclusions.

5.2 THE ARITHMETIC MEAN

5.2.1 Methods of Calculation

You are probably intimately familiar with the arithmetic mean, for whenever you obtain an "average" of grades by summing the grades and dividing by the number of grades, you are calculating the arithmetic mean. In short, *the mean is the sum of the scores or values of a variable divided by their number.* Stated in algebraic form:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_N}{N} = \frac{\sum X}{N}, \tag{5.1}$$

where

$\overline{X}$ = the mean and is referred to as X bar,*
N = the number of scores, and
$\sum$ = the mathematical verb directing us to sum all the measurements.

Thus the arithmetic mean of the scores 8, 12, 15, 19, 24 is $\overline{X} = \frac{78}{5} = 15.60$.

* In Section 1.2 we indicated that italic letters would be employed to represent sample statistics and Greek letters to represent population parameters. The Greek letter μ will be used to represent the population mean.

Table 5.1

Computational procedures for calculating the mean with ungrouped frequency distributions

X	f	fX	
12	1	12	
11	2	22	
10	5	50	$\overline{X} = \dfrac{\sum fX}{N}$
9	4	36	
8	6	48	$\overline{X} = \dfrac{232}{29}$
7	4	28	
6	3	18	$\overline{X} = 8.00$
5	2	10	
4	2	8	
$N = 29$		$\sum fX = 232$	

Obtaining the mean from an ungrouped frequency distribution. You will recall that we constructed a frequency distribution as a means of eliminating the constant repetition of scores that occur with varying frequency in order to permit a single entry in the frequency column to represent the number of times a given score occurs. Thus, in Table 5.1, we know, from the column headed f, that the score of 8 occurred 6 times. In calculating the mean, then, it is not necessary to add 8 six times since we may multiply the score by its frequency and obtain the same value of 48. Since each score is multiplied by its corresponding frequency prior to summing, we may represent the mean for frequency distributions as follows:

$$\overline{X} = \frac{\sum fX}{N}. \tag{5.2}$$

Obtaining the mean from a grouped frequency distribution, raw score method. The calculation of the mean from a grouped frequency distribution involves essentially the same procedures that are employed with ungrouped frequency distributions. To start with, the midpoint of each interval is used to represent all scores within that interval. The midpoint of each interval is multiplied by its corresponding frequency, and the product is summed and divided by N. The procedures employed in calculating the mean from a grouped frequency distribution are demonstrated in Table 5.2.

Table 5.2
Computational procedures for calculating the mean from a grouped frequency distribution

1 Class interval	2 Frequency f	3 Midpoint X	4 Frequency × midpoint fX	
125–129	2	127	254	$\overline{X} = \dfrac{\sum fX}{N}$
120–124	5	122	610	$= \dfrac{10195}{100}$
115–119	8	117	936	$= 101.95$
110–114	10	112	1120	
105–109	15	107	1605	
100–104	20	102	2040	
95–99	15	97	1455	
90–94	10	92	920	
85–89	8	87	696	
80–84	4	82	328	
75–79	3	77	231	
	$N = 100$		$\sum fX = 10195$	

5.2.2 Properties of the Arithmetic Mean

One of the most important properties of the mean is that *it is the point in a distribution of measurements or scores about which the summed deviations are equal to zero.* In other words,

$$\sum(X - \overline{X}) = 0. \tag{5.3}$$

The algebraic proof of this statement is

$$\begin{aligned}\sum(X - \overline{X}) &= \sum X - \sum \overline{X} \\ &= N\overline{X} - N\overline{X} \\ &= 0.\end{aligned}$$

In following this algebraic proof, it is important to note that (1) since

$$\overline{X} = \frac{\sum X}{N},$$

it follows that $\sum X = N\overline{X}$ and (2) summing the mean over all the scores ($\sum \overline{X}$) is the same as multiplying $\overline{X}$ by N, that is, $N\overline{X}$.

Table 5.3

Comparison of the means of two arrays of scores, one of which contains an extreme value

X_1	X_2
2	2
3	3
5	5
7	7
8	33
$\sum X_1 = 25$	$\sum X_2 = 50$
$\overline{X}_1 = 5.00$	$\overline{X}_2 = 10.00$

Therefore the mean is a score or a potential score which balances all the scores on either side of it. In this sense, it is analogous to the fulcrum on a teeter board. In playing on teeter boards, you may have noticed that it is possible for a small individual to balance a heavy individual by moving the latter closer to the fulcrum. Thus, if you wanted to balance a younger brother or sister (assuming they are lighter than you) on a teeter board, you would move yourself toward the center of the board. This analogy leads to a second important characteristic of the mean; that is, *the mean is very sensitive to extreme measurements when these measurements are not balanced on both sides of it.*

Observe the two *arrays* of scores in Table 5.3. An *array is an arrangement of data according to their magnitude from the smallest to the largest value.* Note that all the scores in both distributions are the same except for the very large score of 33 in column X_2. This one extreme score is sufficient to double the size of the mean. The sensitivity of the mean to extreme scores is a characteristic which has important implications governing our use of it. These implications will be discussed in Section 5.5, in which we compare the three measures of central tendency.

A third important characteristic of the mean is that *the sum of squares of deviations from the arithmetic mean is less than the sum of squares of deviations about any other score or potential score.*

To illustrate this characteristic of the mean, Table 5.4 shows the squares and the sum of squares when deviations are taken from the mean and various other scores in a distribution. It can be seen that the sum of squares is smallest in column 2, when deviations are taken from the mean.

This property of the mean provides us with another definition, i.e., *the mean is that measure of central tendency which makes the sum of squared deviations*

Table 5.4
The squares and sum of squares of deviations taken from various scores in a distribution

1 X	2 $(X - \overline{X})^2$	3 $(X - 2)^2$	4 $(X - 3)^2$	5 $(X - 5)^2$	6 $(X - 6)^2$
2	4	0	1	9	16
3	1	1	0	4	9
4	0	4	1	1	4
5	1	9	4	0	1
6	4	16	9	1	0

$N = 5$
$\overline{X} = 4$

around it minimal. The method of locating the mean by finding the minimum sum of squares is referred to as the *least squares* method. The least squares method is of considerable value in statistics, particularly when it is applied to curve fitting.

5.2.3 The Weighted Mean

Let us imagine that four classes in introductory sociology obtained the following mean scores on the final examination: 75, 78, 72, and 80. Could you sum these four means together and divide by four to obtain an overall mean for all four classes? This could be done *only if* the n in each class is identical. What if, as a matter of fact, the mean of 75 is based on an n of 30, the second mean is based on 40 observations, the third on $n = 25$, and the fourth on $n = 50$.

The total sum of scores may be obtained by multiplying each mean by its respective n and summing.

Thus,

$$\begin{aligned}\sum(n \cdot \overline{X}) &= 30(75) + 40(78) + 25(72) + 50(80) \\ &= 11{,}170.\end{aligned}$$

The weighted mean, $\overline{X}_w$, can be expressed as the sum of the mean of each group multiplied by its respective weight (the n in each group) divided by the

sum of the weights (i.e., $\sum w = \sum n = N$).

$$\overline{X}_w = \frac{\sum(w \cdot \overline{X})}{\sum w} = \frac{\sum(n \cdot \overline{X})}{N}$$

$$\begin{aligned}\overline{X}_w &= \frac{30(75) + 40(78) + 25(72) + 50(80)}{145} \\ &= \frac{11{,}170}{145} \\ &= 77.03\end{aligned}$$

5.3 THE MEDIAN

With grouped frequency distributions, the *median* is defined as *that score or potential score in a distribution of scores, above and below which one-half of the frequencies fall.* If this definition sounds vaguely familiar to you, it is not by accident. The median is merely a special case of a percentile rank. Indeed, the median is the score at the 50th percentile. It should be clear that the generalized procedures discussed in Chapter 4 for determining the score at various percentile ranks may be applied to the calculation of the median.

Modifying formula (4.4) for application to the special case of the median, we obtain the following:

$$\text{median} = X_{ll} + i\,\frac{(N/2 - \text{cum}\, f_{ll})}{f_i}. \tag{5.4}$$

Applied to the data appearing in Table 4.1 (Section 4.2.2) the median becomes

$$\begin{aligned}\text{median} &= 109.5 + 5\,\frac{(110/2 - 43)}{17} \\ &= 109.5 + 5\,\frac{(55 - 43)}{17} \\ &= 109.5 + 3.53 = 113.03.\end{aligned}$$

5.3.1 The Median of an Array of Scores

Occasionally, it will be necessary to obtain the median when the N is not sufficient to justify casting the data into the form of a frequency distribution or a grouped frequency distribution. Consider the following array of scores: 5, 19, 37, 39, 45. Note that the scores are arranged in order of magnitude and that N is an odd number. A score of 37 is the median since two scores fall above it and

two scores fall below it.* If N is an even number, the median is the arithmetic mean of the two middle values. The two middle values in the array of scores 8, 26, 35, 43, 47, 73 are 35 and 43. The arithmetic mean of these two values is $(35 + 43)/2$, or 39. Therefore the median is 39.

5.3.2 Characteristics of the Median

An outstanding characteristic of the median is its insensitivity to extreme scores. Consider the following set of scores: 2, 5, 8, 11, 48. The median is 8. This is true, in spite of the fact that the set contains one extreme score of 48. Had the 48 been a score of 97, the median would remain the same. This characteristic of the median makes it valuable for describing central tendency in certain types of distributions in which the mean is an unacceptable measure of central tendency due to its sensitivity to extreme scores. This point will be further elaborated in Section 5.5 when the uses of the three measures of central tendency are discussed.

5.4 THE MODE

Of all measures of central tendency, the mode is the most easily determined since it is obtained by inspection rather than by computation. The mode is simply the score which occurs with greatest frequency. For grouped data, the mode is designated as the midpoint of the interval containing the highest frequency count. In Table 5.2 the mode is a score of 102 since it is the midpoint of the interval (100–104) containing the greatest frequency.

In some distributions, which we shall not consider here, there will be two high points which produce the appearance of two humps, as on a camel's back. Such distributions are referred to as being *bimodal*. A distribution containing more than two humps is referred to as being *multimodal*.

5.5 COMPARISON OF MEAN, MEDIAN, AND MODE

We have seen that the mean is a measure of central tendency in which the *sum* of the deviations on one side of it equals the sum of the deviations on the other side. The median, on the other hand, divides the area under the curve into two

* When working with an array of numbers where N is odd, the definition of the median does not quite hold, i.e., in the example above, in which the median is 37, two scores lie below it and two above it, as opposed to one-half of N. If the score of 37 is regarded as falling one-half on either side of the median, this disparity is reconciled.

equal halves so that the *number* of scores below the median equals the *number* of scores above the median.

In general, the arithmetic mean is the preferred statistic for representing central tendency because of several desirable properties that it possesses. To begin with, the mean is a member of a mathematical system which permits its use in more advanced statistical analyses. We have used deviations from the mean to demonstrate two of its most important characteristics, i.e., the sum of deviations is zero and the sum of squares is minimal. Deviations of scores from the mean provide valuable information about any distribution. We shall be making frequent use of deviation scores throughout the remainder of the text. In contrast, deviation scores from the median and the corresponding squared deviations have only limited applications to more advanced statistical considerations. Another important feature of the mean is that it is the more stable or reliable measure of central tendency. If we were to take repeated samples from a given population, the mean would usually show less fluctuation than either the median or the mode. In other words, the mean generally provides a better estimate of the corresponding population parameter.

On the other hand, there are certain situations in which the median is preferred as the measure of central tendency. When the distribution is symmetrical, the mean and the median are identical. Under these circumstances, the mean should be employed. However, as we have seen, when the distribution is markedly skewed, the mean will provide a misleading estimate of central tendency. In column X_2 of Table 5.3, the mean is 10 even though four of the five scores are below this value. Annual family income is a commonly studied variable in which the median is preferred over the mean since the distribution of this variable is distinctly skewed in the direction of high salaries, with the result that the mean overestimates the income obtained by most families.

The median is also the measure of choice in distributions in which there are *indeterminate values*. To illustrate, when running rats in a maze, there will be occasions when one or more rats will simply not run. Their time scores are, therefore, indeterminate. Their "scores" cannot simply be thrown out since the fact of their not running may be of considerable significance in evaluating the effects of the independent variable. Under these circumstances, the median should be employed as the measure of central tendency.

The mode is the appropriate statistic whenever a quick, rough estimate of central tendency is desired or when we are interested only in the typical case. It is rarely used in the behavioral sciences.

5.6 THE MEAN, MEDIAN, AND SKEWNESS

In Chapter 3, we demonstrated several forms of skewed distributions. We pointed out, however, that skew cannot always be determined by inspection. If you have understood the differences between the mean and the median, you

should be able to suggest a method for determining whether or not a distribution is skewed and, if so, determine the direction of the skew. The basic fact to keep in mind is that the mean is drawn in the direction of the skew whereas the median, unaffected by extreme scores, is not. Thus, when the mean is higher than the median, the distribution may be said to be positively skewed; when the mean is lower than the median, the distribution is negatively skewed. Figure 5.1 demonstrates the relation between the mean and the median in positively and negatively skewed distributions.

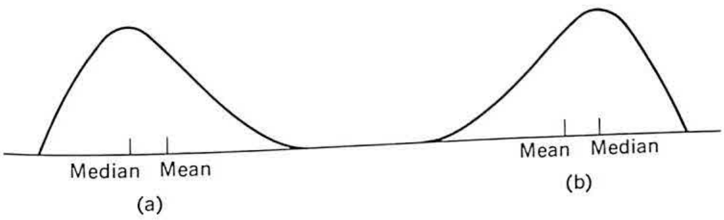


Fig. 5.1 The relation between the mean and median in (a) positively and (b) negatively skewed distributions.

CHAPTER SUMMARY

In this chapter we discussed, demonstrated the calculation of, and compared three indices of central tendency that are frequently employed in the description of frequency distributions: the mean, the median, and the mode.

We saw that the mean may be defined variously as the sum of scores divided by their number, the point in a distribution which makes the summed deviations equal to zero, or the point in the distribution which makes the sum of the squared deviations minimal. The median divides the area of the curve into two equal halves so that the number of scores below the median equals the number of scores above it. Finally, the mode is defined as the most frequently occurring score. We demonstrated the method for obtaining the weighted mean of a set of means when each of the individual means is based on a different n.

Because of special properties it possesses, the mean is the most frequently employed measure of central tendency. However, because of its sensitivity to extreme scores which are not balanced on both sides of the distribution, the median is usually the measure of choice when distributions are markedly skewed. The mode is rarely employed in the behavioral sciences.

Finally, we demonstrated the relationship between the mean and the median in negatively and positively skewed distributions.

Terms to Remember:

Measures of central tendency
Mean
Weighted mean
Median
Mode
Sum of squares
Least squares method
Array

EXERCISES

1. Find the mean, the median, and the mode for each of the following sets of measurements. Show that $\sum(X - \overline{X}) = 0$.
   a) 10, 8, 6, 0, 8, 3, 2, 2, 8, 0
   b) 1, 3, 3, 5, 5, 5, 7, 7, 9
   c) 120, 5, 4, 4, 4, 2, 1, 0
2. In which of the sets of measurements in Problem 1 is the mean a poor measure of central tendency? Why?
3. For each of the sets of measurements in Problem 1, show that *the sum of squares of deviations from the arithmetic mean is less than the sum of squares of deviations about any other score or potential score.*
4. You have calculated the maximum speed of various automobiles. You later discover that all the speedometers were set 5 miles per hour too fast. How will the measures of central tendency based on the corrected data compare with those calculated from the original data?
5. You have calculated measures of central tendency on the weights of barbells, expressing your data in terms of ounces. You decide to recompute after you have divided all the weights by 16 to convert them to pounds. How will this affect the measures of central tendency?
6. Calculate the measures of central tendency for the data in Chapter 3, Problems 13 and 20.
7. In Problem 1(c) above, if the score of 120 were changed to a score of 20, how would the various measures of central tendency be affected?
8. On the basis of the following measures of central tendency, indicate whether or not there is evidence of skew and, if so, what is its direction?

a) $\overline{X} = 56$	Median = 62	Mode = 68	
b) $\overline{X} = 68$	Median = 62	Mode = 56	
c) $\overline{X} = 62$	Median = 62	Mode = 62	
d) $\overline{X} = 62$	Median = 62	Mode = 30,	Mode = 94

9. What is the nature of the distributions in Problem 8(c) and (d)?
10. Calculate the mean of the following array of scores: 3, 4, 5, 5, 6, 7.
    a) Add a constant, say 2, to each score. Recalculate the mean. Generalize: What is the effect on the mean of adding a constant to all scores?
    b) Subtract the same constant from each score. Recalculate the mean. Generalize: What is the effect on the mean of subtracting a constant from all scores?
    c) Alternately add and subtract the same constant, say 2, from the array of scores (that is, 3 + 2, 4 − 2, 5 + 2, etc). Recalculate the mean. Generalize: What is the effect on the mean of adding and subtracting the same constant an equal number of times from an array of scores?
    d) Multiply each score by a constant, say 2. Recalculate the mean. Generalize: What is the effect on the mean of multiplying each score by a constant?
    e) Divide each score by the same constant. Recalculate the mean. Generalize: What is the effect on the mean of dividing each score by a constant?

11. Refer to Problem 18, Chapter 3. Which measure of central tendency might best be used to describe group A? group B? Why?

12. In Section 5.5 we stated that the mean is usually *more reliable* than the median, i.e., less subject to fluctuations. We conducted an experiment consisting of 30 tosses of three dice. The results were as follows: 6, 6, 2; 4, 1, 1; 6, 5, 5; 6, 4, 3; 4, 2, 1; 5, 4, 3; 4, 4, 3; 6, 6, 4; 5, 3, 2; 6, 3, 3; 4, 3, 2; 6, 4, 1; 6, 4, 2; 5, 1, 1; 6, 5, 4; 2, 1, 1; 5, 4, 3; 5, 4, 4; 4, 3, 1; 4, 2, 2; 6, 5, 3; 5, 1, 1; 6, 5, 2; 6, 3, 3; 6, 6, 5; 6, 5, 4; 6, 2, 1; 5, 4, 3; 5, 4, 1; 6, 3, 1.
   a) Calculate the 30 means and 30 medians.
   b) Employing grouping intervals starting with the real limits of the lower interval 0.5–1.5, group the means and medians into separate frequency distributions.
   c) Draw histograms for the two distributions. Do they support the contention that the mean is a more stable estimator of central tendency? Explain.
   d) Given that we were to place the medians and means into two separate hats and were to draw one at random from each. What is the likelihood that:
      i) a statistic greater than 5.5 would be obtained?
      ii) a statistic less than 1.5 would be obtained?
      iii) a statistic greater than 5.5 or less than 1.5 would be obtained?

13. If we know that the mean and median of a set of scores are equal, what can we say about the form of the distribution?

14. During the course of the year you purchased a common stock in the following amounts and prices: 150 at \$3 per share, 400 at \$2.50 per share, 100 at \$4.25 per share, 200 at \$3.50 per share.
   a) What is the break-even point?
   b) If the commissions on the purchases were \$13.50, \$30.00, \$12.75, and \$21.00, respectively, calculate the break-even point, taking the commission into account. (*Hint:* Consider adding the sum of the commissions to $\sum w \cdot X$ prior to dividing).

15. In a departmental final exam, the following mean grades were obtained based on n's of 25, 40, 30, 45, 50, and 20: 72.5, 68.4, 75.0, 71.3, 70.6, and 78.1.
   a) What is the total mean over all sections of the course?
   b) Draw a line graph showing the size of the class along the abscissa and the corresponding means along the ordinate. Does there appear to be a relationship between class size and mean grades on the final exam?

16. Give examples of data in which the preferred measure of central tendency would be the
   a) mean, b) median, c) mode.

17. Find the mean, median, and mode for the following array of scores: −9.0, −6.0, −5.0, −5.0, −0.5, 0, 0.1, 2.0, 4.0, 5.0.

18. Find the mean, median, and mode of the numbers in Problem 17 by first adding 9.0 to each number and then subtracting 9.0 from each of the measures of central tendency.

19. List at least three specific instances in which a measure of central tendency was important in describing a group of people.

20. List at least three specific instances in which a measure of central tendency was utilized to compare groups of people.

21. List at least three specific instances in which a measure of central tendency was estimated for a large group of people from data obtained from a sample of that large group.

22. On the basis of examination performance, an instructor identifies the following groups of students:
    a) Those with a percentile rank of 90 or higher.
    b) Those with a percentile rank of 10 or less.
    c) Those with percentile ranks between 40 and 49.
    d) Those with percentile ranks between 51 and 60.

    Which group would the instructor work with if he wished to raise the *median* performance of the total group? Which group would he work with if he wished to raise the *mean* performance of the total group?

23. Which of the measures of central tendency is most affected by the degree of skew in the distribution? Explain.

24. What can we say about the relationships among the mean, median, and mode in a negatively skewed distribution? in a positively skewed distribution?

Measures of Dispersion 6

6.1 INTRODUCTION

In the introduction to Chapter 4, we saw that a score by itself is meaningless, and takes on meaning only when it is compared with other scores or other statistics. Thus if we know the mean of the distribution of a given variable, we can determine whether a particular score is higher or lower than the mean. But how much higher or lower? It is clear at this point that a measure of central tendency such as the mean only provides a limited amount of information. To more fully describe a distribution, or to more fully interpret a score, it is clear that additional information is required concerning the *dispersion* of scores about our measure of central tendency.

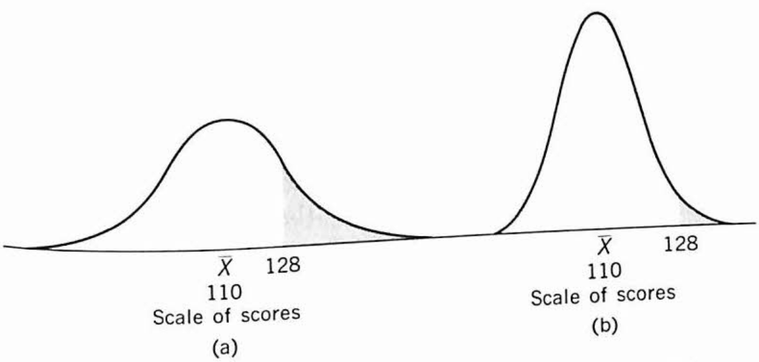


Fig. 6.1 Two frequency polygons with identical means but differing in dispersion or variability.

Consider Fig. 6.1, parts (a) and (b). In both examples of frequency polygons, the mean of the distribution is exactly the same. However, note the difference in the interpretations of a score of 128. In (a), because the scores are

widely dispersed about the mean, a score of 128 may be considered only moderately high. Quite a few individuals in the distribution scored above 128, as indicated by the area to the right of 128. In (b), on the other hand, the scores are compactly distributed about the same mean. This is a more homogeneous distribution. Consequently, the score of 128 is now virtually at the top of the distribution and it may therefore be considered a very high score.

It can be seen, then, that in interpreting individual scores, we must find a companion to the mean or the median. This companion must in some way express the degree of dispersion of scores about the measure of central tendency. We shall discuss five such measures of dispersion or variability: the *range*, the *interquartile range*, the *mean deviation*, the *variance*, and the *standard deviation*. Of the five, we shall find the standard deviation to be our most useful measure of dispersion in both descriptive and inferential statistics. In advanced inferential statistics, as in analysis of variance (Chapter 15), the variance will become a most useful measure of variability.

6.2 THE RANGE

When we calculated the various measures of central tendency, we located a *single point* along the scale of scores and identified it as the mean, the median, or the mode. When our interest shifts to measures of dispersion, however, we must look for an index of variability which indicates the *distance* along the scale of scores.

One of the first measures of distance which comes to mind is the so-called *crude range*. The range is by far the simplest and the most straightforward measure of dispersion. It consists simply of the scale distance between the largest and the smallest score. Indeed, we have already calculated the range in this text although we have not identified it as such. You will recall that, during the course of determining the width of the class interval (Section 3.1.1), we subtracted the lowest score from the highest score and added 1. Thus we have already determined the range of scores in Table 3.1, that is, $(154 - 80) + 1$ or 75.

Although the range is meaningful, it is of little use because of its marked instability. Note that if there is one extreme score in a distribution, the dispersion of scores will appear to be large when, in fact, the removal of that score may reveal an otherwise "compact" distribution. Several years ago, an inmate of an institution for retarded persons was found to have an I.Q. score in the 140's. Imagine the erroneous impression that would result if the range of scores for the inmates was reported as, say, 20–140 or 121! Stated another way, the range reflects only the two most extreme scores in a distribution.

6.3 THE INTERQUARTILE RANGE

In order to overcome the instability of the crude range as a measure of dispersion, the *interquartile range* is sometimes employed. The interquartile range is calculated by simply subtracting the score at the 25th percentile (referred to as the first quartile or Q_1) from the score at the 75th percentile (the third quartile or Q_3). Although this measure of variability of scores is far more meaningful than the crude range, it has two important shortcomings: (1) like the crude range, it does not by itself permit the precise interpretation of a score within a distribution, and (2) like the median, it does not enter into any of the "higher" mathematical relationships that are basic to inferential statistics. Consequently, we shall not devote any more discussion to the interquartile range.

6.4 THE MEAN DEVIATION

In Chapter 5, we pointed out that, when we are dealing with data from normally distributed populations, the mean is our most useful measure of central tendency. We obtained the mean by adding together all the scores and dividing them by N. If we carried these procedures one step further, we could subtract the mean from each score, sum the deviations from the mean, and thereby obtain an estimate of the typical amount of deviation from the mean. By dividing by N, we would have a measure that would be analogous to the arithmetic mean except that it would represent the dispersion of scores from the arithmetic mean.

If you think for a moment about the characteristics of the mean which we discussed in the preceding chapter, you will encounter one serious difficulty. The sum of the deviations of all scores from the mean must add up to zero. Thus if we defined the mean deviation (M.D.) as this sum divided by N, the mean deviation would have to be zero. You will recall that we employed the fact that $\sum(X - \overline{X}) = 0$ to arrive at one of several definitions of the mean.

Now, if we were to add all the deviations *without regard to sign* and divide by N, we would still have a measure reflecting the mean deviation from the arithmetic mean. The resulting statistic would, of course, be based upon the *absolute value* of the deviations. The absolute value of a positive number or of zero is the number itself. The absolute value of a negative number can be found by changing the sign to a positive one. Thus the absolute value of $+3$ or -3 is 3. The symbol for an absolute value is $||$. Thus $|-3| = 3$.

The calculation of the mean deviation is shown in Table 6.1.

As a basis for comparison of the dispersion of several distributions, the mean deviation has some value. For example, the greater the mean deviation, the greater the dispersion of scores. However, for interpreting scores within a

distribution, the mean deviation is less useful since there is no precise mathematical relationship between the mean deviation, as such, and the location of scores within a distribution.

You may wonder why we have bothered to demonstrate the mean deviation when it is of so little use in statistical analysis. There are two reasons: (1) The standard deviation and the variance, which have great value in statistical analysis, are very close relatives of the mean deviation. In order to calculate the standard deviation and the variance, we shall need only to add one column to Table 6.1. (2) We want you to understand that measures of dispersion represent, in a sense, bases for estimating errors in prediction.

Table 6.1

Computational procedures for calculating the M.D. from an array of scores

X	$(\lvert X - \overline{X}\rvert)$	
9	$\lvert +4\rvert$	$$\text{M.D.} = \frac{\sum(\lvert X - \overline{X}\rvert)^*}{N} \qquad (6.1)$$
8	$\lvert +3\rvert$	
7	$\lvert +2\rvert$	
7	$\lvert +2\rvert$	
7	$\lvert +2\rvert$	$\text{M.D.} = \frac{26}{15} = 1.73$
5	$\lvert 0\rvert$	
5	$\lvert 0\rvert$	
5	$\lvert 0\rvert$	
5	$\lvert 0\rvert$	
4	$\lvert -1\rvert$	
4	$\lvert -1\rvert$	
3	$\lvert -2\rvert$	
3	$\lvert -2\rvert$	
2	$\lvert -3\rvert$	
1	$\lvert -4\rvert$	

$\sum X = 75$ $\sum(\lvert X - \overline{X}\rvert) = 26$
$N = 15$
$\overline{X} = 5$

* For scores arranged in the form of a frequency distribution, the following formula for the mean deviation should be used:

$$\text{M.D.} = \sum f(\lvert X - \overline{X}\rvert)/N.$$

Before taking up the standard deviation, let us examine the second point a bit further by posing a question. In the absence of any specific information, what is our best single basis for predicting a score that is obtained by any given individual? If the data are drawn from a normally distributed population, it turns out that the mean (or *any* measure of central tendency) is our best single predictor. The largest error that we can make in prediction when we employ the mean is the most extreme score minus the mean. On the other hand, if we use *any* other score for prediction, the maximum error possible is the entire range of the distribution. This large an error will occur when we predict the highest score in the distribution for an individual actually obtaining the lowest score, or vice versa.

In our subsequent discussion of the standard deviation and other statistics based on it, we shall stress the fact that these measures provide estimates of error in the prediction of scores.

6.5 THE VARIANCE (s^2) AND STANDARD DEVIATION (s)*

Following a perusal of Table 6.1, you might pose this question, "We had to treat the values in the column headed $(X - \overline{X})$ as absolute numbers because their sum was equal to zero. Why could we not square each $(X - \overline{X})$, and then add the squared deviations? In this way, we would rid ourselves of the minus signs in a perfectly legitimate way while still preserving the information that is inherent in these deviation scores."

The answer is: We could, if, by so doing, we arrived at a statistic which is of greater value in judging dispersion than those we have already discussed. It is most fortunate that the standard deviation, based on the squaring of these deviation scores, is of immense value in three different respects. (1) The standard deviation reflects dispersion of scores so that the variability of different distributions may be compared in terms of the standard deviation (s). (2) The standard deviation permits the *precise* interpretation of scores within a distribution. (3) The standard deviation, like the mean, is a member of a *mathematical system* which permits its use in more advanced statistical considerations. Thus, we shall employ measures based upon s when we advance into inferential statistics. We shall have more to say about the interpretive aspects of s after we have shown how it is calculated.

---

* We remind you that italic letters are used to represent sample statistics, and Greek letters to represent population parameters; e.g., σ^2 represents the population variance and σ represents the population standard deviation. The problem of estimating population parameters from sample values will be discussed in Chapter 12.

6.5.1 Calculation of variance and standard deviation, mean deviation method, with ungrouped scores

The *variance* is defined verbally as *the sum of the squared deviations from the mean divided by N*. Symbolically, it is represented as

$$s^2 = \frac{\sum(X - \overline{X})^2}{N}. \tag{6.2}\dagger$$

At times it is more convenient to use the symbol x to represent $X - \overline{X}$. Thus

$$x = X - \overline{X}. \tag{6.3}$$

The variance then becomes

$$s^2 = \frac{\sum x^2}{N}. \tag{6.4}$$

The standard deviation is the square root of the variance and is defined as

$$s = \sqrt{\frac{\sum(X - \overline{X})^2}{N}}, \tag{6.5}$$

or

$$s = \sqrt{\frac{\sum x^2}{N}}. \tag{6.6}$$

The computational procedures for calculating the standard deviation, utilizing the mean deviation method, are shown in Table 6.2.

You will recall that the sum of the $(X - \overline{X})^2$ column [that is, $\sum(X - \overline{X})^2$] is known as the *sum of squares* or the *sum squares* and that this sum is minimal when deviations are taken about the mean. From this point on in the course we shall encounter the sum of squares with regularity. It will take on a number of different forms, depending on the procedures that we elect for calculating it. However, it is important to remember that, whatever the form, the sum squares represents the *sum of the squared deviations from the mean.*

The mean deviation method was shown only to impress you with the fact that the standard deviation is based on the deviation of scores from the mean. This method is extremely unwieldy for use in calculation, particularly when the mean is a fractional value, which is usually the case. Consequently, in the succeeding paragraphs we shall examine a number of alternative ways of calculating the sum squares.

---

† The important distinction between biased $\sum(X - \overline{X})^2/N$ and unbiased

$$\sum(X - \overline{X})^2/(N - 1)$$

estimates of the population variance will be discussed in Chapter 12.

Table 6.2

Computational procedures for calculating s, mean deviation method, from an array of scores

X	$X - \overline{X}$	$(X - \overline{X})^2$	
9	+4	16	$s = \sqrt{\frac{\sum(X - \overline{X})^2}{N}}$*
8	+3	9	$= \sqrt{72/15}$
7	+2	4	$= \sqrt{4.80}$
7	+2	4	$= 2.19$
7	+2	4	
5	0	0	
5	0	0	
5	0	0	
5	0	0	
4	−1	1	
4	−1	1	
3	−2	4	
3	−2	4	
2	−3	9	
1	−4	16	
$\sum X = 75$	$\sum(X - \overline{X}) = 0$	$\sum(X - \overline{X})^2 = 72$	
$N = 15$			
$\overline{X} = 5$			

* For data cast in the form of a frequency distribution, the formula for the standard deviation is

$$s = \sqrt{\frac{\sum f(X - \overline{X})^2}{N}}.$$

Note that the f appears in the formula to remind you that each $(X - \overline{X})^2$ should be multiplied by its corresponding frequency prior to summing. Even when we are dealing with an array of scores, this formula is the most general for the frequency of each score is one, i.e., $f = 1$. For this reason, you should regard the f as implied even when it is not given.

6.5.2 Calculation of Standard Deviation, Raw Score Method, with Ungrouped Scores

It can be shown mathematically that

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{N}, \tag{6.7}$$

where

$$\sum x^2 = \sum(X - \overline{X})^2 = \sum X^2 - 2\sum X\overline{X} + \sum \overline{X}^2.$$

Table 6.3

Computational procedures for calculating s, raw score method, from an array of scores

X	X^2	
1	1	$s = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2}$
2	4	$= \sqrt{\frac{447}{15} - 5^2}$
3	9	$= \sqrt{29.80 - 25.00}$
3	9	$= \sqrt{4.80} = 2.19$
4	16	
4	16	
5	25	
5	25	
5	25	
5	25	
7	49	
7	49	
7	49	
8	64	
9	81	
$\sum X = 75$ $N = 15$ $\overline{X} = 5$	$\sum X^2 = 447$	

However, $\sum X = N\overline{X}$ and summing the mean square over all values of $\overline{X}$ is the same as multiplying by N. Thus

$$\begin{aligned}\sum x^2 &= \sum X^2 - 2N\overline{X}^2 + N\overline{X}^2 \\ &= \sum X^2 - N\overline{X}^2 \\ &= \sum X^2 - N\frac{(\sum X)^2}{N^2} \\ &= \sum X^2 - \frac{(\sum X)^2}{N}.\end{aligned}$$

This leads naturally to the development of a useful raw score formula for s:

$$s = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2}\,; \tag{6.8}$$

thus

$$s = \sqrt{\frac{\sum x^2}{N}} = \sqrt{\frac{\sum X^2 - (\sum X)^2/N}{N}}$$

$$= \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\frac{\sum X^2}{N} - \overline{X}^2}\,.$$

You will note that the result agrees with the answer which we obtained by the mean deviation method. Table 6.3 summarizes the computational procedure.

6.5.3 Errors to Watch For

In calculating the standard deviation, using the raw score method, it is common for students to confuse the similar-appearing terms $\sum X^2$ and $(\sum X)^2$. It is important to remember that the former represents the sum of the squares of each of the individual scores, whereas the latter represents the square of the sum of the scores. By definition, it is impossible to obtain a negative sum of squares or a negative standard deviation. In the event that you obtain a negative value under the square root sign, you have probably confused these two terms.

A rule of thumb for estimating the standard deviation is that the ratio of the range to the standard deviation is rarely smaller than 2 or greater than 6. In our preceding example, the ratio is $9/2.19 = 4.11$. If we obtain a standard deviation which yields a ratio greater than 6 or smaller than 2, we have almost certainly made an error.

6.6 INTERPRETATION OF THE STANDARD DEVIATION

An understanding of the meaning of the standard deviation hinges on a knowledge of the relationship between the standard deviation and the normal distribution. Thus, in order to be able to interpret the standard deviations that are calculated in this chapter, it will be necessary to explore the relationship between the raw scores, the standard deviation, and the normal distribution. This material is presented in the following chapter.

CHAPTER SUMMARY

We have seen that to fully describe a distribution of scores, we require more than a measure of central tendency. We must be able to describe how these scores are dispersed about central tendency. In this connection we discussed

five measures of dispersion: the range, the interquartile range, the mean deviation, the standard deviation, and the variance.

For normally distributed variables, the two measures based on the squaring of deviations about the mean (the variance and the standard deviation) are maximally useful. We discussed and demonstrated the procedures for calculating the standard deviation employing the mean deviation method and the raw score method with ungrouped frequency distributions. We also pointed out several of the errors commonly made in calculating standard deviations.

Terms to Remember:

Dispersion
Range
Interquartile range
Mean deviation
Mean deviation method
Raw score method
Standard deviation
Variance
Absolute value of a number
Sum of squares

EXERCISES

1. Calculate s^2 and s for the following array of scores: 3, 4, 5, 5, 6, 7.
   a) Add a constant, say, 2, to each score. Recalculate s^2 and s. Would the results be any different if you had added a larger constant, say, 200? Generalize: What is the effect on s and s^2 of adding a constant to an array of scores? Does the variability increase as we increase the magnitude of the scores?
   b) Subtract the same constant from each score. Recalculate s^2 and s. Would the results be any different if you had subtracted a larger constant, say, 200? Generalize: What is the effect on s and s^2 of subtracting a constant from an array of scores?
   c) Alternately add and subtract the same constant from each score (i.e., $3 + 2$, $4 - 2$, $5 + 2$, etc). Recalculate s and s^2. Would the results be any different if you had added and subtracted a larger constant? Generalize: What is the effect on s and s^2 of adding and subtracting a constant from an array of scores? (*Note:* this generalization is extremely important with relation to subsequent chapters when we discuss the effect of random errors on measures of variability.)
   d) Multiply each score by a constant, say, 2. Recalculate s and s^2. Generalize: What is the effect on s and s^2 of multiplying each score by a constant?
   e) Divide each score by the same constant. Recalculate s and s^2. Generalize: What is the effect on s and s^2 of dividing each score by a constant?
2. Compare these above generalizations with those made in relation to the mean (see Problem 10, Chapter 5).
3. A rigorous definition of a measure of variation as a descriptive statistic would involve the following properties: (a) If a constant is added or subtracted from

each score or observation, the measure of variation remains unchanged. (b) If each score is multiplied or divided by a constant, the measure of variation is also multiplied or divided by that number. Check the following for the satisfaction of these conditions:

i) the mean,	ii) the median,	iii) the mode,
iv) the mean deviation,	v) the standard deviation,	vi) the variance.

If the properties defining measures of dispersion were extended to include *powers* of the constant by which each score is multiplied, would the variance qualify as a measure of dispersion?

4. How would the standard deviation be affected by the situations described in Problems 4 and 5, Chapter 5?
5. What is the nature of the distribution if $s = 0$?
6. Calculate the standard deviations for the following sets of measurements.

   a) 10, 8, 6, 0, 8, 3, 2, 2, 8, 0
   b) 1, 3, 3, 5, 5, 5, 7, 7, 9
   c) 20, 1, 2, 5, 4, 4, 4, 0
   d) 5, 5, 5, 5, 5, 5, 5, 5, 5, 5

7. Why is the standard deviation in part (c) of Problem 6 so large? Describe the effect of extreme deviations on s.
8. Determine the range for the sets of measurements in Problem 6. For which of these is the range a misleading index of variability, and why?
9. Calculate the mean and standard deviation for the 250 stocks listed in Problem 7, (Table 3.5), Chapter 3, employing two different grouping intervals. The calculated standard deviation has a somewhat different value in each instance. To what do you attribute the difference?
10. Calculate the mean and standard deviation for the frequency distributions obtained in Problems 8, 9, 10, and 11 of Chapter 3. Compare and discuss the results.
11. A comparison shopper compares prices of chopped chuck at a number of different supermarkets. She finds the following prices per pound (in cents):
    56, 65, 48, 73, 59, 72, 63, 65, 60, 63, 44, 79, 63, 61, 66, 69, 64, 71, 58, 63.

    a) Find the mean.
    b) Find the range, interquartile range, and mean deviation.
    c) Find the standard deviation and variance.

12. Give one advantage of the standard deviation over the variance. Give an example.
13. Referring back to Problem 24, Chapter 3, find the mean and standard deviation of the number of quarts of milk sold at the supermarket.
14. List at least three specific instances in which a measure of variability was important in describing a group.
15. List at least three specific instances in which a measure of variability was important in comparing groups of people.

16. The following data list the maximum daily temperature recorded for New York City for the months of January and May, 1965 and 1966.*

Date	Jan. 1965	Jan. 1966	May 1965	May 1966
1	35	62	71	66
2	29	52	77	60
3	32	46	71	68
4	39	45	90	59
5	43	53	62	70
6	44	47	78	83
7	44	47	52	66
8	49	44	60	61
9	55	30	78	52
10	40	48	92	57
11	35	42	88	63
12	36	27	78	57
13	41	32	80	67
14	37	42	74	64
15	19	41	82	74
16	17	26	87	71
17	16	39	83	68
18	25	38	73	66
19	25	35	70	64
20	38	44	80	84
21	36	42	72	85
22	50	39	80	75
23	47	37	79	85
24	32	38	70	80
25	38	32	79	76
26	44	26	94	84
27	42	36	92	82
28	38	19	83	73
29	27	23	68	82
30	20	38	67	73
31	30	28	74	68

For each month and year:

a) Find the mean.

b) Find the range and mean deviation.

c) Find the standard deviation and variance.

---

* The data for this problem were extracted from *The World Almanac*, pp. 558–559. New York: Newspaper Enterprise Association, Inc., 1967. Reprinted by permission.

The Standard Deviation and the Standard Normal Distribution

7

7.1 INTRODUCTION

We previously noted that scores derived from scales employed by behavioral scientists are generally meaningless by themselves. To take on meaning they must be compared to the distribution of scores from some reference group. Indeed, the scores derived from any scale, including those employed by the physical scientists, become more meaningful when they are compared to some reference group of objects or persons. Thus, if we were to learn that a Canadian fisherman caught a northern pike weighing 50 lb, we might or might not be impressed, depending upon the extent of our knowledge concerning the usual weight of this type of fish. However, once a reference group is established, the measurement becomes meaningful. Since most northern pike weigh under 10 lb and only rarely achieve weights as high as 20 lb, the achievement of our apocryphal fisherman must be considered Bunyanesque.

7.2 THE CONCEPT OF STANDARD SCORES

In interpreting a single score, we want to place it in some position with respect to a collection of scores from some reference group. In Chapter 4, you learned to place a score by determining its percentile rank. It will be recalled that the percentile rank of a score tells us the percentage of scores that are of lower scale value. Another approach for interpretation of a single score might be to view it with reference to some central point, such as the mean. Thus, a score of 20 in a distribution with a mean of 23 might be reported as -3. Finally, we might express this deviation score in terms of standard deviation units. Thus, if our standard deviation is 1.5, the score of 20 would be two standard deviations below the mean (that is, $-3/1.5 = -2$). This process of dividing a deviation of a score from the mean by the standard deviation is known as the transformation

to z-scores. Symbolically, z is defined as

$$z = \frac{X - \overline{X}}{s}. \tag{7.1}$$

Since we previously employed x to represent $(X - \overline{X})$, we may also state that

$$z = x/s. \tag{7.2}$$

It will be noted that every score in the distribution may be transformed into a z-score, in which case each z will represent the *deviation of a specific score from the mean expressed in standard deviation units.*

In order to fully appreciate the value of transforming to z-scores, let us look at some of the properties of z-scores.

1. The sum of the z-scores is zero. Symbolically stated,

$$\sum z = 0. \tag{7.3}$$

2. The mean of z-scores is zero. Thus

$$\bar{z} = \frac{\sum z}{N} = 0. \tag{7.4}$$

3.* The sum of the squared z-scores equals N. Thus

$$\sum z^2 = N. \tag{7.5}$$

This characteristic may be demonstrated mathematically:

$$\begin{aligned}\sum z^2 &= \frac{\sum(X - \overline{X})^2}{s^2} = \frac{1}{s^2} \cdot \sum(X - \overline{X})^2 \\ &= \frac{N}{\sum(X - \overline{X})^2} \cdot \sum(X - \overline{X})^2 \\ &= N\end{aligned}$$

4. The standard deviation and the variance of z-scores is one. Thus

$$\sigma z = \sigma z^2 = 1. \tag{7.6}$$

To demonstrate:

$$\sigma_z^2 = \frac{\sum(z - \bar{z})^2}{N}.$$

* This property of z-scores is important for understanding one of many alternative formulas for calculating the Pearson correlation coefficient r (See Section 8.2.)

Since $\bar{z} = 0$, then

$$\sigma_z^2 = \frac{\sum z^2}{N}.$$

Since $\sum z^2 = N$, then

$$\sigma_z^2 = \frac{N}{N} = 1.$$

What is the value of transforming to a z-score? Simply this: if the *population of scores* on a given variable is normal, we may express any score as a percentile rank by referring our z to the *standard normal distribution.* In addition, since z-scores represent abstract numbers as opposed to the concrete values of the original scores (inches, pounds, I.Q. scores, etc.), we may compare an individual's position on one variable with his position on a second. To understand these two important characteristics of z-scores, we must make reference to the *standard normal distribution.*

7.3 THE STANDARD NORMAL DISTRIBUTION

The standard normal distribution has a μ of 0, a σ of 1, and a total area equal to 1.00.* There is a fixed proportion of cases between a vertical line, or ordinate erected at any one point and an ordinate erected at any other point. Taking a few reference points along the normal curve, we can make the following statements:

1) Between the mean and 1 standard deviation above the mean are found 34.13% of all cases. Similarly, 34.13% of all cases fall between the mean and 1

---

* It will be recalled that the Greek letters μ and σ represent the population mean and the standard deviation, respectively. The equation of the normal curve is:

$$Y = \frac{Ni}{\sigma\sqrt{2\pi}} e^{\frac{-(X-\mu)^2}{2\sigma^2}}$$

in which

Y = the frequency at a given value of X,
σ = the standard deviation of the distribution,
π = a constant equaling approximately 3.1416,
e = approximately 2.7183,
N = total frequency of the distribution,
μ = the mean of the distribution,
i = the width of the interval,
X = any score in the distribution.

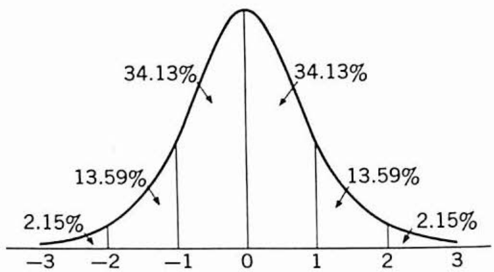
It should be clear that there are a family of curves which may be called normal. By setting $Ni = 1$, a distribution is generated in which $\mu = 0$ and total area under the curve equals 1.

standard deviation below the mean. Stated in another way, 34.13% of the *area* under the curve is found between the mean and 1 standard deviation above the mean, and 34.13% of the *area* falls between the mean and -1 standard deviation.

2) Between the mean and 2 standard deviations above the mean are found 47.72% of all cases. Since the normal curve is symmetrical, 47.72% of the area also falls between the mean and -2 standard deviations.

3) Finally, between the mean and 3 standard deviations above the mean are found 49.87% of all the cases. Similarly, 49.87% of the cases fall between the mean and -3 standard deviations. Thus, 99.74% of all cases fall between ± 3 standard deviations. These relationships are shown in Fig. 7.1.

Fig. 7.1 Areas between selected points under the normal curve.



Now, by transforming the scores of a normally distributed variable to z-scores, we are, in effect, expressing these scores in units of the standard normal curve. For any given value of X with a certain proportion of area beyond it, there is a corresponding value of z with the same proportion of area beyond it. Thus, if we have a population in which $\mu = 30$ and $\sigma = 10$, the z of a score at the mean ($X = 30$) will equal zero, and the z of scores 1 standard deviation above and below the mean ($X = 40$ and $X = 20$) will be $+1.00$ and -1.00, respectively.

7.3.1 Finding Area Between Given Scores

For expositional purposes, we confined our preceding discussion of area under the standard normal curve to selected points. As a matter of actual fact, however, it is possible to determine the percent of areas between *any* two points by making use of the tabled values of the area under the normal curve (Table A). The left-hand column headed by z represents the deviation from the mean expressed in standard deviation units. *By referring to the body of the table, we can determine the proportion of total area between a given score and the mean* (Column B), *and the area beyond a given score* (Column C). Thus, if an individual obtained a score of 24.65 on a normally distributed variable with $\mu = 16$ and $\sigma = 5$,

his z-score would be

$$z = \frac{24.65 - 16}{5} = 1.73.$$

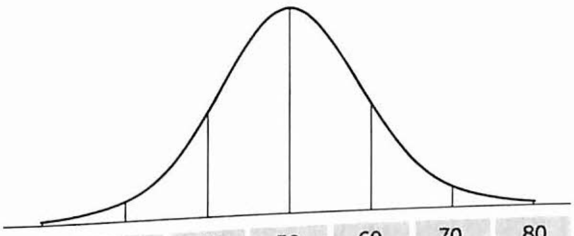
Referring to Column B in Table A, we find that 0.4582 or 45.82%* of the area lies between his score and the mean. Since 50% of the area also falls below the mean in a symmetrical distribution, we may conclude that 95.82% of all the area falls below a score of 24.65. Note that we can now translate this score into a percentile rank of 95.82.

Let us suppose another individual obtained a score of 7.35 on the same normally distributed variable. His z-score would be

$$z = \frac{7.35 - 16}{5} = -1.73.$$

Since the normal curve is symmetrical, only the areas corresponding to the positive z-values are given in Table A. Negative z-values will have precisely the same proportions as their positive counterparts. Thus, the area between the mean and a z of -1.73 is also 45.82%. The percentile rank of a score below the mean may be obtained either by subtracting 45.82% from 50%, or directly from Column C. In either case, the percentile rank of a score of 7.35 is 4.18.

You should carefully note that these relationships apply *only to scores from normally distributed populations.* Transforming the raw scores to standard scores does not, in any way, alter the form of the original distribution. The only change is to convert the mean to zero and the standard deviation to one.



Raw score	20	30	40	50	60	70	80
z-score	−3	−2	−1	0	+1	+2	+3
Percentile rank	.13	2.28	15.87	50.00	84.13	97.72	99.87

Fig. 7.2 Relationships among raw scores, z-scores, and percentile ranks of a normally distributed variable in which $\mu = 50$ and $\sigma = 10$.

* The areas under the normal curve are expressed as proportions of area. To convert to percentage of area, multiply by 100 or merely move the decimal two places to the right.

Thus, if the original distribution of scores is nonnormal, *the distribution of z*-scores will be nonnormal. In other words, our transformation to z's will *not* convert a nonnormal distribution to a normal distribution.

Figure 7.2 further clarifies the relationships among raw scores, z-scores, and percentile ranks of a normally distributed variable. It assumes that $\mu = 50$ and $\sigma = 10$.

7.4 ILLUSTRATIVE PROBLEMS

Let us take several sample problems in which we assume that the mean of the general population, μ, is equal to 100 on a standard I.Q. test, and the standard deviation, σ, is 16. It is assumed that the variable is normally distributed.

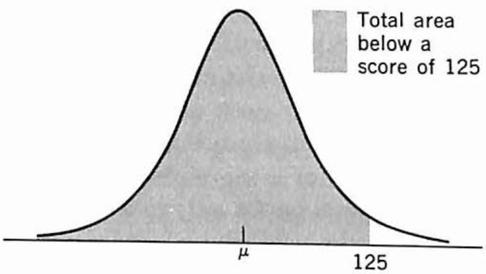


Fig. 7.3 Proportion of area below a score of 125 in a normal distribution with $\mu = 100$ and $\sigma = 16$.

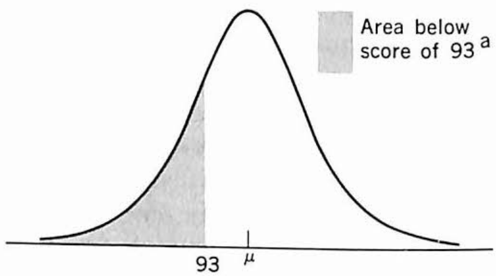


Fig. 7.4 Proportion of area below a score of 93 in a normal distribution with $\mu = 100$ and $\sigma = 16$.

Problem 1

John Doe obtains a score of 125 on an I.Q. test. What percent of cases fall between his score and the mean? What is his percentile rank in the general population?

At the outset, it is wise to construct a crude diagram representing the relationships in question. Thus, in the present example, the diagram would appear as shown in Fig. 7.3. To find the value of z corresponding to $X = 125$, we subtract the population mean from 125 and divide by 16. Thus

$$z = \frac{125 - 100}{16} = 1.56.$$

Looking up 1.56 in Column B (Table A), we find that 44.06% of the area falls between the mean and 1.56 standard deviations above the mean. John Doe's percentile rank is, therefore, 50 + 44.06 or 94.06.

Problem 2

Mary Jones scores 93 on an I.Q. test. What is her percentile rank in the general population (Fig. 7.4)?

$$z = \frac{93 - 100}{16} = -0.44.$$

The minus sign indicates that the score is below the mean. Looking up 0.44 in Column C, we find that 33.00% of the cases fall below her score. Thus her percentile rank is 33.00.

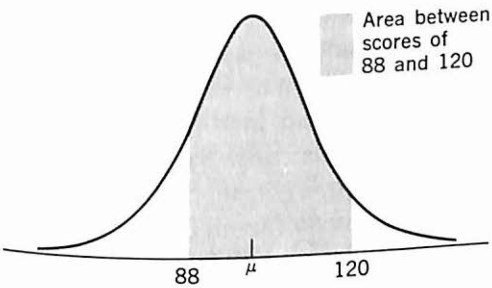


Fig. 7.5 Proportion of area between the scores 88 and 120 in a normal distribution with $\mu = 100$ and $\sigma = 16$.

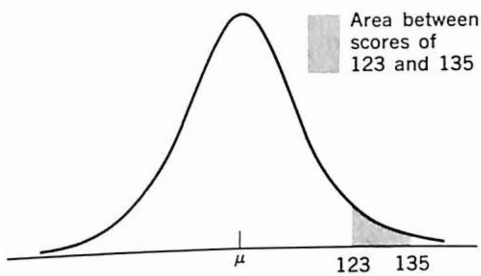


Fig. 7.6 Proportion of area between the scores 123 and 135 in a normal distribution with $\mu = 100$ and $\sigma = 16$.

Problem 3

What percent of cases fall between a score of 120 and a score of 88 (Fig. 7.5)?

Note that to answer this question we do *not* subtract 88 from 120 and divide by σ. The areas in the normal probability curve are designated in relation to the mean as a fixed point of reference. We must, therefore, separately calculate the area between the mean and a score of 120 and the area between the mean and a score of 88. We then add the two areas to answer our problem.

Procedure:

Step 1. Find the z corresponding to $X = 120$:

$$z = \frac{120 - 100}{16} = 1.25.$$

Step 2. Find the z corresponding to $X = 88$:

$$z = \frac{88 - 100}{16} = -0.75.$$

Step 3. Find the required areas by referring to Column B (Table A):

Area between the mean and $z = 1.25$ is 39.44%.
Area between the mean and $z = -0.75$ is 27.34%.

Step 4. Add the two areas together.

Thus, the area between 88 and 120 = 66.78%.

Problem 4

What percent of the area falls between a score of 123 and 135 (Fig. 7.6)?

Again, we cannot obtain the answer directly, we must find the area between the mean and a score of 123 and subtract this from the area between the mean and a score of 135.

Procedure:

Step 1. Find the z corresponding to $X = 135$.

$$z = \frac{135 - 100}{16} = 2.19.$$

Step 2. Find the z corresponding to $X = 123$.

$$z = \frac{123 - 100}{16} = 1.44.$$

Step 3. Find the required areas by referring to Column B.

Area between the mean and $z = 2.19$ is 48.57%.

Area between the mean and $z = 1.44$ is 42.51%.

Step 4. Subtract to obtain the area between 123 and 135. The result is

$$48.57 - 42.51 = 6.06\%.$$

Problem 5

We stated earlier that our transformation to z-scores permits us to compare an individual's position on one variable with his position on another. Let us illustrate this important use of z-scores.

On a standard aptitude test, John G. obtained a score of 245 on the verbal scale and 175 on the mathematics scale. The means and the standard deviations of each of these normally distributed scales are as follows: Verbal, $\mu = 220$, $\sigma = 50$; Math, $\mu = 150$, $\sigma = 25$. On which scale did John score higher? All that we need to do is compare John's z-score on each variable. Thus:

$$\text{Verbal } z = \frac{245 - 220}{50} = 0.50 \qquad \text{Math } z = \frac{175 - 150}{25} = 1.00.$$

We may conclude, therefore, that John scored higher on the math scale of the aptitude test. Of course, if we so desire, we may express these scores as percentile ranks. Thus John's percentile rank is 84.13 on the math scale and only 69.15 on the verbal scale.

7.5 THE STANDARD DEVIATION AS AN ESTIMATE OF ERROR

When discussing the mean deviation at the beginning of Chapter 6, we pointed out that the mean of a distribution can be considered our best single predictor of a score in the absence of any other information. The more compactly our scores are distributed about the mean, the smaller our errors will be in prediction, on the average. Since the standard deviation reflects the dispersion of scores, it becomes, in a sense, an estimate of error. Thus, if we have two distributions with identical means but with standard deviations of, say, 10 and 30 respectively, we will make larger errors, on the average, when we employ the mean as a basis for predicting scores in the latter distribution. This characteristic of the standard deviation is important in the understanding of certain derivatives of σ which we shall be discussing in subsequent chapters.

7.6 THE TRANSFORMATION TO *T*-SCORES

Many psychological and educational tests have been purposely constructed so as to yield a normal distribution of z-scores. Since it is inconvenient and sometimes confusing to deal with a distribution containing many negative values, the z-scores are frequently converted to T-scores employing the following transformation equation:

$$T^* = 50 + 10z.$$

This transformation now yields a distribution with a mean equal to 50 and a standard deviation equal to 10.

T-scores may readily be reconverted to units of the standard normal curve:

$$z = \frac{T - \overline{T}}{10}.$$

Thus, a person obtaining a T of 65 would have a corresponding

$$z = \frac{65\text{–}50}{10} = 1.50.$$

---

* The transformation may involve the substitution of any desired constants into the equation. Thus, if a mean of 100 and a standard deviation of 20 is desired, the transformation equation becomes: $100 + 20z$.

Employing the standard normal curve, his score is found to have a corresponding percentile rank of 93.32.

CHAPTER SUMMARY

In this chapter, we demonstrated the value of the standard deviation for comparison of the dispersion of scores in different distributions of a variable, the interpretation of a score with respect to a single distribution, and the comparison of scores on two or more variables. We showed how to convert raw scores into units of the standard normal curve (transformation to z-scores) and the various characteristics of the standard normal curve were explained. A series of worked problems demonstrated the various applications of the conversion of normally distributed variables to z-scores.

Finally, we discussed the standard deviation as an estimate of error and as an estimate of precision. We demonstrated the use of the T-transformation as a convenient method for eliminating the negative values occurring when scores are expressed in terms of z.

Terms to Remember:

Standard normal distribution — *Standard score (z-score)*
T-transformation

EXERCISES

1. Given a normal distribution with a mean of 45.2 and a standard deviation of 10.4, find the standard score equivalents for the following scores.
   a) 55 b) 41 c) 45.2
   d) 31.5 e) 68.4 f) 18.9

2. Find the proportion of area under the normal curve between the mean and the following z-scores.
   a) -2.05 b) -1.90 c) -0.25
   d) $+0.40$ e) $+1.65$ f) $+1.96$
   g) $+2.33$ h) $+2.58$ i) $+3.08$

3. Given a normal distribution based on 1000 cases with a mean of 50 and a standard deviation of 10, find
   a) the proportion of area and the number of cases between the mean and the following scores:

   60, 70, 45, 25.

b) the proportion of area and the number of cases *above* the following scores:

60, 70, 45, 25, 50.

c) the proportion of area and the number of cases between the following scores:

60 — 70, 25 — 60, 45 — 70, 25 — 45.

4. Below are given Student Spiegel's scores, the mean, and the standard deviation on each of three tests given to 3,000 students.

Test	μ	σ	Spiegel's score
Arithmetic	47.2	4.8	53
Verbal comprehension	64.6	8.3	71
Geography	75.4	11.7	72

a) Convert each of Spiegel's test scores to standard scores.
b) On which test did Spiegel stand highest? On which lowest?
c) Spiegel's score in arithmetic was surpassed by how many students? In Verbal Comprehension? In Geography?
d) What assumption must be made in order to answer the preceding question?

5. On a normally distributed mathematics aptitude test, for females,

$$\mu = 60, \qquad \sigma = 10,$$

and for males,

$$\mu = 64, \qquad \sigma = 8.$$

a) Arthur obtained a score of 62. What is his percentile rank on both the male and the female norms?
b) Helen's percentile rank is 73 on the female norms. What is her percentile rank on the male norms?

6. If frequency polygons were constructed for each of the following, which do you feel would approximate a normal curve?
a) Heights of a large representative sample of adult American males.
b) Means of a large number of samples with a fixed N (say, $N = 100$) drawn from a normally distributed population of scores.
c) Means of a large number of samples of a fixed N (say, $N = 100$), drawn from a moderately skewed distribution of scores.
d) Weights, in ounces, of ears of corn selected randomly from a cornfield.
e) Annual income, in dollars, of the "breadwinner" of a large number of American families selected at random.
f) Weight, in ounces, of all fish caught in a popular fishing resort.

7. In a normal distribution with $\mu = 72$ and $\sigma = 12$:
a) What is the score at the 25th percentile?
b) What is the score at the 75th percentile?
c) What is the score at the 90th percentile?

d) Find the percent of cases scoring above 80.
e) Find the percent of cases scoring below 66.
f) Between what scores do the middle 50 percent of the cases lie?
g) Beyond what scores do the most extreme 5 percent lie?
h) Beyond what scores do the most extreme 1 percent lie?

8. Answer the above questions (a through h) for

$$\mu = 72 \quad \text{and} \quad \sigma = 8;$$
$$\mu = 72 \quad \text{and} \quad \sigma = 4;$$
$$\mu = 72 \quad \text{and} \quad \sigma = 2.$$

9. Employing the following information, did Larry do better on Test I or Test II? Which test did Mindy do better on?

	I	*II*
Larry	18	20
Mindy	17	22
Alan	17	22
Gary	16	21
John	12	31

10. Are all sets of z-scores normally distributed? Why?
11. Is there more than one normal distribution?
12. Employing the data in Problem 16, Chapter 6, convert each of the temperature readings for the four time periods to standard scores.
13. The following data list the number of home runs made by the home run leaders in the National and American Leagues from 1943–1966:

Year	National League	American League
1943	29	34
1944	33	22
1945	28	24
1946	23	44
1947	51	32
1948	40	39
1949	54	43
1950	47	37
1951	42	33
1952	37	32
1953	47	43
1954	49	32
1955	51	37
1956	43	52
1957	44	42
1958	47	42

Year	National League	American League
1959	46	42
1960	41	40
1961	46	61
1962	49	48
1963	44	45
1964	47	49
1965	52	32
1966	44	49

a) Find the mean and standard deviation for each league.
b) Convert the number of home runs for each league to standard scores.

Correlation 8

8.1 THE CONCEPT OF CORRELATION

Up to this point in the course, we have been interested in calculating various statistics which permit us to thoroughly describe the distribution of the values of a single variable, and to relate these statistics to the interpretation of individual scores. However, as you are well aware, many of the problems in the behavioral sciences go beyond the description of a single variable in its various and sundry ramifications. We are frequently called upon to determine the relationships among two or more variables. For example, college administration officers are vitally concerned with the relationship between high school averages or College Entrance Examination Board (CEEB) scores and performance at college. Do students who do well in high school or who score high on the CEEB also perform well in college? Conversely, do poor high school students or those who score low on the CEEB perform poorly at college? Do parents with high intelligence tend to have children of high intelligence? Is there a relationship between the declared dividend on stocks and their paper value in the exchange? Is there a relationship between the socio-economic class and recidivism in crime?

As soon as we raise questions concerning the relationships among variables, we are thrust into the fascinating area of correlation. In order to express quantitatively the extent to which two variables are related, it is necessary to calculate a *correlation coefficient*. There are many types of correlation coefficients. The decision of which one to employ with a specific set of data depends upon such factors as (1) the type of scale of measurement in which each variable is expressed; (2) the nature of the underlying distribution (continuous or discrete); and (3) the characteristics of the distribution of the scores (linear or nonlinear). We present two correlation coefficients in this text: the Pearson r, or the *Pearson product moment correlation coefficient*, employed with interval or ratio scaled variables, and r_{rho} or the *Spearman rank order correlation coefficient* employed with ordered or ranked data.

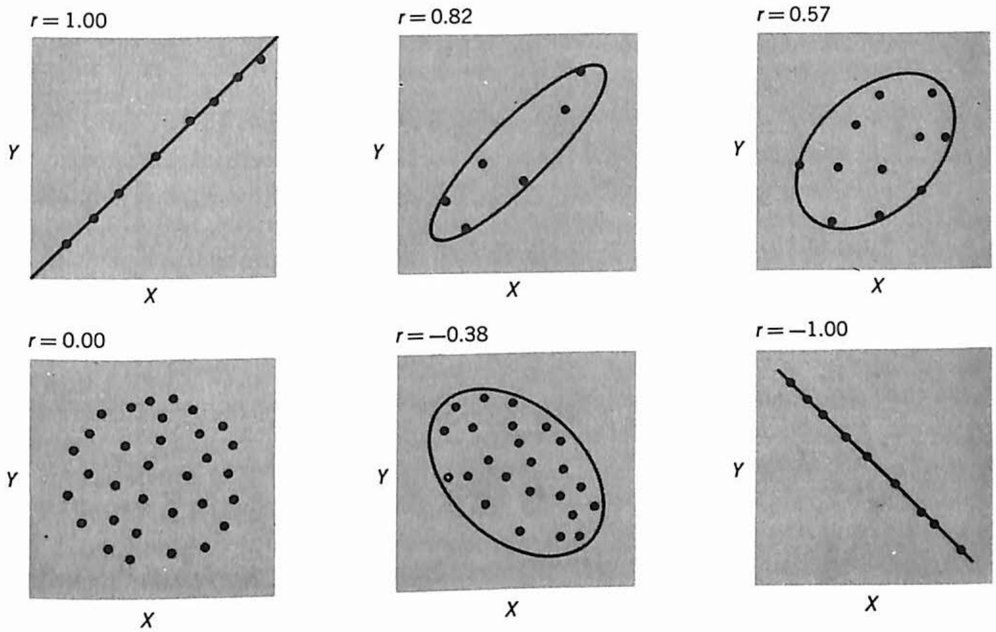


Fig. 8.1 Scatter diagrams showing various degrees of relationship between two variables.

No matter which correlational technique we use, all have certain characteristics in common.

1. Two sets of measurements are obtained on the same individuals (or events) or on pairs of individuals who are matched on some basis.

2. The values of the correlation coefficients vary between +1.00 and −1.00. Both of these extremes represent perfect relationships between the variables, and 0.00 represents the absence of a relationship.

3. A *positive relationship* means that individuals obtaining high scores on one variable tend to obtain high scores on a second variable. The converse is also true, i.e., individuals scoring low on one variable tend to score low on a second variable.

4. A *negative relationship* means that individuals scoring low on one variable tend to score high on a second variable. Conversely, individuals scoring high on one variable tend to score low on a second variable.

Figure 8.1 shows a series of *scatter diagrams* illustrating various degrees of relationships between two variables, X and Y. In interpreting the figures it is important to remember that every circle represents two values: an individual's score on the X-variable and the same person's score on the Y-variable. As indicated earlier (Section 3.4), the X-variable is represented along the abscissa and the Y-variable along the ordinate.

Table 8.1

Raw scores and corresponding z-scores made by 7 subjects on two variables (hypothetical data)

Subject	X	x	x^2	z_x	Y	y	y^2	z_y	$z_x z_y$
A	1	−6	36	−1.5	4	9	81	−1.5	2.25
B	3	−4	16	−1.0	7	6	36	−1.0	1.00
C	5	−2	4	−0.5	10	3	9	−0.5	0.25
D	7	0	0	0	13	0	0	0	0
E	9	2	4	0.5	16	3	9	0.5	0.25
F	11	4	16	1.0	19	6	36	1.0	1.00
G	13	6	36	1.5	22	9	81	1.5	2.25

$\sum X = 49 \quad \sum x^2 = 112 \quad \sum Y = 91 \quad \sum y^2 = 252 \quad \sum z_x z_y = 7.00$

$\overline{X} = 7.00 \quad s_x = \sqrt{\frac{112}{7}} = 4.00 \quad \overline{Y} = 13.00 \quad s_y = \sqrt{\frac{252}{7}} = 6.00$

8.2 PEARSON r AND z-SCORES

A high positive Pearson r indicates that each individual obtains approximately the same z-score on both variables. In a perfect positive correlation, each individual obtains *exactly* the same z-score on both variables.

With a high negative r, each individual obtains approximately the same z-score on both variables, but the sigma scores are opposite in sign.

Remembering that the z-score represents a measure of relative position on a given variable (i.e., a high positive z represents a high score relative to the remainder of the distribution, and a high negative z represents a low score relative to the remainder of the distribution), we may now generalize the meaning of the Pearson r.

Pearson r represents the extent to which the same individuals or events occupy the same relative position on two variables.

In order to explore the fundamental characteristics of the Pearson r, let us examine a simplified example of a perfect positive correlation. In Table 8.1, we find the paired scores of 7 individuals on the two variables, X and Y.

It will be noted that the scale values of X and Y do not need to be the same for the calculation of a Pearson r. In the example, we see that X ranges from 1 through 13, whereas Y ranges from 4 through 22. This independence of r from specific scale values permits us to investigate the relationships among an unlimited variety of variables. We can even correlate the length of the big toe with the I.Q. if we feel so inclined!

Note, also, as we have already pointed out, that the z-scores of each subject on each variable are identical in the event of a perfect positive correlation. Had we reversed the order of either variable, i.e., paired 1 with 22, paired 3 with 19,

etc., the z-scores would still be identical, but would be opposite in sign. In this latter case, our correlation would be a maximum negative ($r = -1.00$).

If we multiply our paired z-scores and then sum the results, we will obtain maximum values only when our correlation is 1.00. Indeed, as the correlation approaches zero, the sum of the products of the paired z-scores also approaches zero. Note that when the correlation is perfect, the sum of the products of the paired z-scores is equal to n, where n equals the number of pairs. These facts lead to one of the many different but algebraically equivalent formulas for r.

$$r = \frac{\sum(z_x z_y)}{n}. \tag{8.1*}$$

It is suggested that you take the data in Table 8.1, rearrange them in a number of different ways, and calculate r, employing the above formula. You will arrive at a far more thorough understanding of r in this way than by reading the text (not that we are discouraging the latter).

It so happens that the formula is unwieldy in practice since it requires the calculation of separate z's for each score of each individual. Imagine the Herculean task of calculating r when n exceeds 50 cases, as it often does in behavioral research!

For this reason, a number of different computational formulas are employed. In this text, we shall illustrate the use of two: (1) the mean deviation formula, and (2) the raw score formula.

8.3 CALCULATION OF PEARSON r

8.3.1 Mean Deviation Method

The mean deviation method for calculating a Pearson r, like the z-score formula above, is not often employed by behavioral scientists because it involves more time and effort than other computational techniques. It is being presented here primarily because it sheds further light on the characteristics of the Pearson r. However, with small n's, it is as convenient a computational formula as any, unless an automatic calculator is available. The computational formula for the Pearson r, employing the mean deviation method is

$$r = \frac{\sum xy \quad \text{(cross products)}}{\sqrt{\sum x^2 \cdot \sum y^2}}. \tag{8.2}$$

* In Section 7.2, we pointed out that $\sum z^2 = N$. You will note that when the correlation is perfect, each z-score on the X-variable is identical to its corresponding z-score on the Y-variable. Thus $\sum z_x z_y = \sum z_x^2 = \sum z_y^2$, when $r = 1.00$. In other words, in a perfect correlation, $\sum z_x z_y = n$. The Pearson r then becomes n/n or 1.00.

Let us illustrate the mean deviation method employing the figures in Table 8.1 but arranged in a different sequence (Table 8.2).

The computational procedures, employing the mean deviation method, should be perfectly familiar to you. The $\sum x^2$ and the $\sum y^2$ have already been confronted when we were studying the standard deviation. In fact, in calculating r only one step has been added, namely, the one to obtain the sum of the cross products ($\sum xy$). This is obtained easily enough by multiplying the deviation of each individual's score from the mean of the X-variable by his corresponding deviation on the Y-variable and then summing all of the cross products. Incidentally, you should notice the similarity of $\sum xy$ to $\sum(z_x z_y)$ which is discussed in Section 8.2. Everything that has been said with respect to the relationship between the variations in $\sum(z_x z_y)$ and r holds also for $\sum xy$ and r. Notice that, if maximum deviations in X had lined up with maximum deviations in Y, and so on down through the array, $\sum xy$ would have been equal to 168.00, which is the same as the value of the denominator, and would have produced a correlation of 1.00.

8.3.2 Raw Score Method

We have already seen that the raw score formula for calculating the sum of squares is

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{N} \qquad \text{and} \qquad \sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{N}.$$

By analogy, the raw score formula for the sum of the cross products is

$$\sum xy = \sum XY - \frac{(\sum X)(\sum Y)}{n}. \tag{8.3}$$

In calculating the Pearson r, by the raw score method, you have the option of calculating all the above quantities separately and substituting them into formula (8.2) or defining r in terms of raw scores as in formula (8.4) or (8.5) as follows:

$$r = \frac{\sum XY - \dfrac{(\sum X)(\sum Y)}{n}}{\sqrt{\left[\sum X^2 - \dfrac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \dfrac{(\sum Y)^2}{n}\right]}} \tag{8.4}$$

or

$$r = \frac{\dfrac{\sum XY}{n} - \overline{X}\overline{Y}}{\sqrt{\left(\dfrac{\sum X^2}{n} - \overline{X}^2\right)\left(\dfrac{\sum Y^2}{n} - \overline{Y}^2\right)}}. \tag{8.5}$$

Table 8.2

Computational procedures for Pearson *r* employing mean deviation method (hypothetical data)

Subject	X	x	x^2	Y	y	y^2	xy
A	1	−6	36	7	−6	36	36
B	3	−4	16	4	−9	81	36
C	5	−2	4	13	0	0	0
D	7	0	0	16	3	9	0
E	9	2	4	10	−3	9	−6
F	11	4	16	22	9	81	36
G	13	6	36	19	6	36	36

$\sum x^2 = 112$ $\quad\sum y^2 = 252 \quad \sum xy = 138$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}} = \frac{138}{\sqrt{(112)(252)}} = \frac{138}{168.00} = 0.82$$

Table 8.3

Computational procedures for Pearson *r* employing raw score method (hypothetical data)

Subject	X	X^2	Y	Y^2	XY
A	1	1	7	49	7
B	3	9	4	16	12
C	5	25	13	169	65
D	7	49	16	256	112
E	9	81	10	100	90
F	11	121	22	484	242
G	13	169	19	361	247

$\sum X = 49 \quad \sum X^2 = 455 \quad \sum Y = 91 \quad \sum Y^2 = 1435 \quad \sum XY = 775$

$\overline{X} = 7 \quad \overline{Y} = 13$

$$s_x = \sqrt{\tfrac{455}{7} - (7)^2} = \sqrt{65 - 49} = \sqrt{16} = 4$$

$$s_y = \sqrt{\tfrac{1435}{7} - (13)^2} = \sqrt{205 - 169} = \sqrt{36} = 6$$

$$r = \frac{\dfrac{\sum XY}{n} - \overline{X}\overline{Y}}{s_x s_y} = \frac{\frac{775}{7} - (7)(13)}{(4)(6)}$$

$$= \frac{110.71 - 91}{24} = \frac{19.71}{24} = 0.82$$

You may have noticed that the denominator in formula (8.5) consists of the standard deviation of X (s_x) multiplied by the standard deviation of Y (s_y). This provides an alternative formula for calculating the Pearson r, namely

$$r = \frac{\frac{\sum XY}{n} - \overline{X}\overline{Y}}{s_x s_y}. \tag{8.6}$$

The procedures for calculating r by the raw score method are summarized in Table 8.3. Here we find exactly the same coefficient as we did before. As with the mean deviation method, all the procedures, except those of obtaining the cross products, are familiar to you from our earlier use of the raw score formula to obtain the standard deviation. The quantity $\sum XY$ is obtained very simply by multiplying each X-value by its corresponding Y and then summing these products.

8.4 A WORD OF CAUTION

When low correlations are found, one is strongly tempted to conclude that there is little or no relationship between the two variables under study. However, it must be remembered that the Pearson r reflects only the *linear* relationship between two variables. The failure to find evidence of a relationship may be due to one of two possibilities: (1) the variables are, in fact, unrelated, or (2) the variables are related in a *nonlinear* fashion. In the latter instance, the Pearson r would not be an appropriate measure of the degree of relationship between the variables. To illustrate, if we were plotting the relationship between age and strength of grip, we might obtain a picture somewhat like Fig. 8.2.

It is usually possible to determine whether there is a substantial departure from linearity by examining the scatter diagram. If the distribution of points in the scatter diagram is elliptical, without the decided bending of the ellipse that occurs in Fig. 8.2, it may safely be assumed that the relationship is linear. Any small departures from linearity will not greatly influence the size of the correlation coefficient.

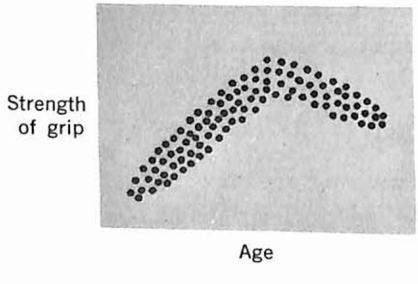


Fig. 8.2 Scatter diagram of two variables which are related in a nonlinear fashion (hypothetical data).

On the other hand, where there is marked curvilinearity, as in Fig. 8.2, a curvilinear coefficient of correlation would better reflect the relationship between the two variables under investigation. Although it is beyond the scope of this text to investigate nonlinear coefficients of correlation, you should be aware of this possibility and, as a matter of course, you should construct a scatter diagram prior to your calculation of the Pearson r.

The assumption of linearity of relationship is the most important requirement to justify the use of the Pearson r as a measure of relationship between two variables. It is not necessary that r be calculated only with normally distributed variables. So long as the distributions are unimodal and relatively symmetrical, a Pearson r may legitimately be computed.

Another situation giving rise to spuriously low correlation coefficients results from restricting the range of values of one of the variables. For example, if we were interested in the relationship between age and height for children from 3 years to 16 years of age, undoubtedly we would obtain a rather high coefficient of correlation between these two variables. However, suppose that we were to restrict the range of one of our variables? What effect would this have on the size of the coefficient? That is, let us look at the same relationship between age and height but only for those children between the ages of 9 and 10. We would probably end up with a rather low coefficient. Let us look at this graphically.

You will note that the overall relationship illustrated in Fig. 8.3 is rather high. The inset illustrates what happens when we restrict our range. Note that the scatter diagram contained in the inset represents a very low correlation. This restriction of the range is frequently referred to as the *truncated range*. The problem of truncated range is not uncommon in behavioral research, since much of this research is conducted in the colleges and universities where subjects have been preselected for intelligence and related variables. Thus they represent a fairly homogeneous group with respect to these variables. Consequently, when an attempt is made to demonstrate the relationship between

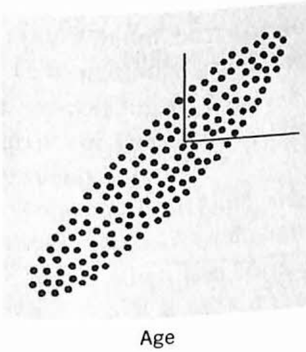


Fig. 8.3 Scatter diagram illustrating high correlation over entire range of X- and Y-values, but low correlation when range is truncated (hypothetical data).

variables like CEEB scores and college grades, the resulting coefficient may be lowered because of the truncated range. Furthermore, the correlations would be expected to be lower for colleges which select their students from within a narrow range of CEEB scores.

8.5 ORDINALLY SCALED VARIABLES AND r_{rho}

Let us imagine that you are a grade school teacher. After long years of observation in the classroom, you have developed a strong suspicion that intelligence and leadership are related variables. In an effort to test this hypothesis, you obtain I.Q. estimates on all the children in your class. However, you discover that no scales are available to measure classroom leadership and you can think of no satisfactory way to quantify this variable. Nevertheless, from numerous observations of the children in different leadership situations, you feel confident that you are able to rank the children from those highest in leadership to those lowest in this quality. The resulting measurements constitute, of

Table 8.4

Computational procedures for calculating r_{rho} from ranked variables (hypothetical data)

I.Q. rank	Leadership rank	D	D^2
1	4	−3	9
2	2	0	0
3	9	−6	36
4	1	3	9
5	7	−2	4
6	10	−4	16
7	8	−1	1
8	13	−5	25
9	5	4	16
10	3	7	49
11	11	0	0
12	6	6	36
13	12	1	1
14	15	−1	1
15	14	1	1
		$\sum D = 0$	$\sum D^2 = 204$

$$\begin{aligned} r_{\text{rho}} &= 1 - \frac{6\sum D^2}{n(n^2-1)} \\ &= 1 - \frac{6 \cdot 204}{15(224)} \\ &= 1 - \frac{1224}{3360} \\ &= 1 - 0.36 \\ &= 0.64 \end{aligned}$$

course, an ordinal scale. Although we could obtain a Pearson r with ranked data, a variant of the product moment correlation coefficient, which is referred to as the Spearman r, r_{rho}, or the rank correlation coefficient, reduces the computational task involved in obtaining the correlation. The Spearman r is appropriate when one scale constitutes ordinal measurement and the remaining scale is either ordinal or higher. However, prior to applying the r_{rho} formula, both scales must be expressed as ranks.

Realizing that your knowledge of the children's I.Q. scores might "contaminate" your estimates of their leadership qualities, you ask a fellow teacher to provide ranks for his children based on leadership qualities. You then obtain, independent of ranks, an estimate of their I.Q.'s. Following this, you rank the I.Q.'s from highest to lowest.

The rank correlation coefficient requires that you obtain the differences in the ranks, square and sum the squared differences, and substitute the resulting values into the formula

$$r_{\text{rho}} = 1 - \frac{6\sum D^2}{n(n^2 - 1)}, \tag{8.7}$$

in which $D = \text{rank } X - \text{rank } Y$.

Table 8.4 shows the hypothetical data and the procedures involved in calculating r_{rho}.

As a matter of course, $\sum D$ should be obtained even though it is not used in any of the calculations. It constitutes a useful check on the accuracy of your calculations up to this point since $\sum D$ must equal zero. If you obtain any value other than zero, you should recheck your original ranks and the subsequent subtractions.

Occasionally, when it is necessary to convert scores to ranks, as in the present example, you will find two or more tied scores. In this event, assign the mean rank to each of the tied scores. The next score in the array receives the rank normally assigned to it. Thus, the ranks of the scores 128, 122, 115, 115, 115, 107, 103 would be 1, 2, 4, 4, 4, 6, 7.

You may wonder how the formula for rho reflects the degree of relationship between two ordinally scaled variables. To engage in extensive mathematical proofs is, as you know, beyond the scope of this book. However, a few of the fundamentals of the mathematics involved may permit you to grasp intuitively "how the formula works."

First, you may note that, when the correlation is maximum positive (+1.00), the difference between each pair of ranks is equal to zero; thus $\sum D^2 = 0$. On the other hand, when the correlation is maximum negative (−1.00), the difference between ranks is maximum and $\sum D^2$ is maximum. Now, if we wanted to

"invent" a formula which reflects these facts, we could define rho as follows:

$$r_{\text{rho}} = 1 - \frac{2\sum D^2}{\sum D_{\text{max}}{}^2}.$$

For a moment, let us concentrate on one term in our "invented" formula, that is, $\sum D^2/\sum D^2_{\text{max}}$. This term reflects, in any data we are analyzing, the *obtained proportion of the maximum possible sum of the squared differences* in ranks. If our correlation is maximum positive, in which $\sum D^2 = 0$, the term reduces to zero. If we were to multiply it by any constant, it would, of course, remain zero. On the other hand, if our correlation is maximum negative, the term reduces to unity, i.e., $\sum D^2_{\text{max}}/\sum D^2_{\text{max}} = 1$. In order that the correlation coefficient may vary between $+1.00$ and -1.00, it is necessary to multiply the numerator by the constant 2 and subtract the entire term from 1.00.

Now, it can be proved mathematically that $\sum D^2_{\text{max}} = n(n^2 - 1)/3$. You may check this empirically by setting up several sets of data with maximum negative correlations. You will find that the above formulation always holds.

Substituting this term in our "invented" formula, rho becomes

$$r_{\text{rho}} = 1 - \frac{3(2\sum D^2)}{n(n^2 - 1)} = 1 - \frac{6\sum D^2}{n(n^2 - 1)},$$

which is, of course, the formula for the Spearman rank coefficient.

CHAPTER SUMMARY

In this chapter we discussed the concept of correlation and demonstrated the calculation of two correlation coefficients, i.e., the Pearson r, employed with interval or ratio scaled data and r_{rho} used with ordinally scaled variables.

We saw that correlation is concerned with determining the extent to which two variables are related or tend to vary together. The quantitative expression of the extent of the relationship is given in terms of the magnitude of the correlation coefficient. Correlation coefficients vary between values of -1.00 to $+1.00$; both extremes represent perfect relationships. A coefficient of zero indicates the absence of a relationship between two variables.

We noted that the Pearson r is appropriate only for variables which are related in a linear fashion. With ranked data, the Spearman rank correlation coefficient is the exact counterpart of the Pearson r. The various computational formulas for the Pearson r may be employed in calculating r_{rho} from ranked data. However, a computational formula for r_{rho} was demonstrated which considerably simplifies the calculation of the rank correlation coefficient.

Terms to Remember:

Correlation
Correlation coefficient
Pearson r (product moment correlation coefficient)
Spearman r (r_{rho} or rank correlation coefficient)
Positive relationship
Negative relationship
Scatter diagram
Cross products
Truncated range

EXERCISES

1. The data below show the scores obtained by a group of 20 students on a college entrance examination and a verbal comprehension test. Prepare a scatter diagram and calculate a Pearson r for these data.

Student	College entrance exam (X)	Verbal comprehension test (Y)	Student	College entrance exam (X)	Verbal comprehension test (Y)
A	52	49	K	64	53
B	49	49	L	28	17
C	26	17	M	49	40
D	28	34	N	43	41
E	63	52	O	30	15
F	44	41	P	65	50
G	70	45	Q	35	28
H	32	32	R	60	55
I	49	29	S	49	37
J	51	49	T	66	50

2. The data below represent scores obtained by 10 students on a statistics examination and their final grade point average. Prepare a scatter diagram and calculate a Pearson r for these data.

Student	Statistics examination	Grade point average	Student	Statistics examination	Grade point average
A	90	2.50	F	70	1.00
B	85	2.00	G	70	1.00
C	80	2.50	H	60	0.50
D	75	2.00	I	60	0.50
E	70	1.50	J	50	0.50

3. A psychological study involved the rating of rats along a dominance-submissiveness continuum. In order to determine the reliability of the ratings, the ranks given by two different observers were tabulated:

Animal	Rank observer A	Rank observer B	Animal	Rank observer A	Rank observer B
A	12	15	I	6	5
B	2	1	J	9	9
C	3	7	K	7	6
D	1	4	L	10	12
E	4	2	M	15	13
F	5	3	N	8	8
G	14	11	O	13	14
H	11	10	P	16	16

Are the ratings reliable? Explain your answer.

4. Explain in *your own words* the meaning of correlation.
5. In each of the examples presented below, identify a possible source of contamination in the collection and/or interpretation of the results of a correlational analysis.
   a) The relationship between age and reaction time for subjects from three months to 65 years of age.
   b) The correlation between I.Q. and grades for honor students at a university.
   c) The relationship between vocabulary and reading speed among children in a "culturally deprived" community.
6. For a group of 50 individuals $\sum z_x z_y$ is 41.3. What is the correlation between the two variables?
7. The following scores were made by 5 students on two tests. Calculate the Pearson r (using $r = \sum z_x z_y / n$). Convert to ranks and calculate r_{rho}.

Student	Test X	Test Y
A	5	1
B	5	3
C	5	5
D	5	7
E	5	9

Generalize: What is the effect of tied ranks on r_{rho}?

8. Show algebraically that

$$\sum xy = \sum XY - \frac{\sum X \sum Y}{n}.$$

9. What effect does a departure from linearity have on the Pearson r?
10. How does the range of scores sampled affect the size of the correlation coefficient?
11. In Problem 15, Chapter 5, we showed the mean grades obtained by classes with varying numbers of students. What is the correlation between class size and mean grades on the final exam?
12. Following are the data showing scores on college entrance examinations and college grade-point averages following the first semester. What is the relationship between these two variables?

Entrance examinations	Grade point averages	Entrance examinations	Grade point averages
440	1.57	528	2.08
448	1.83	550	2.15
455	2.05	582	3.44
460	1.14	569	3.05
473	2.73	585	3.19
485	1.65	593	3.42
489	2.02	620	3.87
500	2.98	650	3.00
512	1.79	690	3.12
518	2.63		

13. Following are data showing the latitude of 35 cities in the northern hemisphere and the mean high and mean low annual temperatures.
    a) What is the correlation between latitude and mean high temperature?
    b) What is the correlation between latitude and mean low temperature?
    c) What is the correlation between mean high and mean low temperature?

City	Latitude to nearest degree	Mean high temperature	Mean low temperature
Acapulco	17	88	73
Accra	6	86	74
Algiers	37	76	71
Amsterdam	52	54	46
Belgrade	45	62	45
Berlin	53	55	40
Bogota	5	66	50
Bombay	19	87	74
Bucharest	44	62	42
Calcutta	22	89	70
Casablanca	34	72	55
Copenhagen	56	52	41

(cont.)

City	Latitude to nearest degree	Mean High temperature	Mean low temperature
Dakar	15	84	70
Dublin	53	56	42
Helsinki	60	46	35
Hong Kong	22	77	68
Istanbul	41	64	50
Jerusalem	32	74	54
Karachi	25	87	70
Leningrad	60	46	33
Lisbon	39	67	55
London	52	58	44
Madrid	40	66	47
Manila	15	89	73
Monrovia	6	84	73
Montreal	46	50	35
Oslo	60	50	36
Ottawa	45	51	32
Paris	49	59	43
Pnom Penh	12	89	74
Prague	50	54	42
Rangoon	17	89	73
Rome	42	71	51
Saigon	11	90	74
Shanghai	31	69	53

14. Obtain rank correlation coefficients for the data in Problem 13. How closely do the r_{rho}'s approach the Pearson r's?

15. Refer back to Problem 16, Chapter 6:
    a) Find the correlation between the maximum daily temperature in January, 1965 and January, 1966.
    b) Find the correlation between the maximum daily temperature in May, 1965 and May, 1966.
    c) Is there a correlation between the day of the month and the maximum daily temperature in January, 1965? January, 1966? May, 1965? May, 1966?

16. Obtain rank correlation coefficients for Problems 15(a) and (b). Compare the r_{rho}'s with the previously obtained Pearson r's.

17. Listed below are the 1967 earnings per share of 37 United States Industries and the closing cost per share of stock for February 15, 1968.
    a) Find the correlation between cost per share and 1967 earnings per share.
    b) Express the earnings as a percentage of cost per share.
    c) Is there a correlation between percentage of earnings and cost per share?

1967 Earnings per share (in cents)	Cost per share (to nearest dollar)
130	34
56	23
8	7
9	6
200	48
195	22
177	28
7	20
361	46
167	22
93	21
94	14
168	27
86	41
70	19
287	41
214	37
84	28
237	29
33	15
36	14
157	40
304	30
115	8
68	44
73	11
69	34
16	35
88	17
122	16
147	42
160	40
367	50
35	18
38[illegible]	45
159	34
105	12

18. Employ the data in Problem 13, Chapter 7.
    a) Determine whether there is any relationship between the number of home runs obtained by the National and American League leaders over the 24-year period.
    b) Assuming that the year constitutes an ordinal scale, determine the correlation between the year and the number of home runs hit by the leader in each league.
19. Explain the difference between $r = 0.76$ and $r = -0.76$.

Regression and Prediction 9

9.1 INTRODUCTION TO PREDICTION

Knowing a person's I.Q., what can we say about his prospects of satisfactorily completing a college curriculum? Knowing his prior voting record, can we make any informed guesses concerning his vote in the coming election? Knowing his mathematics aptitude score, can we estimate the quality of his performance in a course in statistics?

Let us look at an example. Suppose we are trying to predict Student Jones' score on the final exam. If the only information available was that the class mean on the final was 75 ($\overline{Y} = 75$), the best guess we could make is that he would obtain a score of 75 on the final.* However, far more information is usually available, e.g., Mr. Jones obtained a score of 62 on the midterm examination. How can we use this information to make a better prediction about his performance on the final exam? If we know that the class mean on the midterm examination was 70 ($\overline{X} = 70$), we could reason that since he scored below the mean on the midterm, he would probably score below the mean on the final. At this point, we appear to be closing in on a more accurate prediction of his performance. How might we further improve the accuracy of our prediction? Simply knowing that he scored below the mean on the midterm does not give us a clear picture of his relative standing on this exam. If, however, we know the standard deviation on the midterm, we could express his score in terms of his relative position, i.e., his z-score. Let us imagine that the standard deviation on the midterm was 4 ($s_x = 4$). Since he scored 2 standard deviations below the mean ($z_x = -2$), would we be justified in guessing that he would score 2 standard deviations below the mean on the final ($z_y = -2$)? That is, if $s_y = 8$, would you predict a score of 59 on the final? No! You will note that an important piece of information is missing, i.e., the correlation between the

* See Section 5.2.2, in which we demonstrated that the sum of the deviations from the mean is zero and that the sum of squares of deviations from the arithmetic mean is less than the sum of squares of deviations about any other score or potential score.

midterm and the final. You may recall from our discussion of correlation* that the Pearson r represents the extent to which the same individuals or events occupy the same relative position on two variables. Thus, we are only justified in predicting a score of exactly 59 on the final when the correlation is perfect (that is, when $r = +1.00$). Suppose that the correlation is equal to zero. Then it should certainly be obvious that we are not justified in predicting a score of 59; rather, we are once again back to our original prediction of 75 (that is, $\overline{Y}$).

In summary, when $r = 0$, our best prediction is 75 ($\overline{Y}$); when $r = +1.00$, our best prediction is 59 ($z_y = z_x$). It should be clear that predictions from intermediate values of r will fall somewhere between 59 and 75†.

An outstanding advantage, then, of a correlational analysis stems from its application to problems involving predictions from one variable to another. Psychologists, educators, biologists, sociologists, and economists are constantly being called upon to perform this function. To provide an adequate explanation of r and to illustrate its specific applications, it is necessary to digress into an analysis of linear regression.

9.2 LINEAR REGRESSION

To simplify our discussion, let us start with an example of two variables which are usually perfectly or almost perfectly related: monthly salary and yearly income. In Table 9.1 we have listed the monthly income of eight wage earners

Table 9.1
Monthly salaries and annual income of eight wage earners in an electronics firm (hypothetical data)

Employee	Monthly salary	Annual income
A	400	4800
B	450	5400
C	500	6000
D	575	6900
E	600	7200
F	625	7500
G	650	7800
H	675	8100

* See Section 8.2.

† We are assuming that the correlation is positive. If the correlation were -1.00, our best prediction would be a score of 91, that is, $z_y = -z_x$.

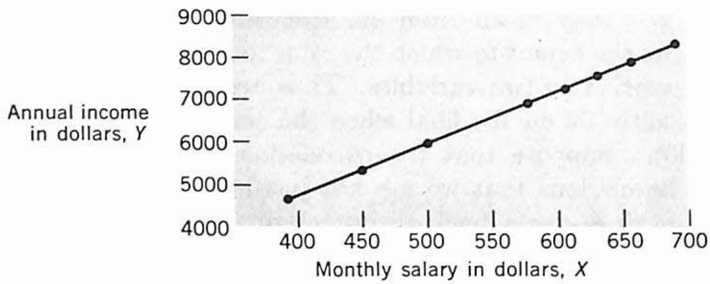


Fig. 9.1 Relation of monthly salaries to annual income for eight employees in an electronics firm.

in a small electronics firm. These data are shown graphically in Fig. 9.1. It is customary to refer to the horizontal axis as the X-axis, or abscissa, and to the vertical axis as the ordinate, or Y-axis. If the variables are temporally related, the prior one is represented on the X-axis. It will be noted that all salaries are represented on a straight line extending diagonally from the lower left-hand corner to the upper right-hand corner.

9.2.1 Formula for Linear Relationships

The formula relating monthly salary to annual salary may be represented as

$$Y = 12X.$$

You may substitute any value of X into the formula and obtain directly the value of Y. For example, if another employee's monthly salary were \$700, his annual income would be

$$Y = 12 \cdot 700 = 8400.$$

Let us add one more factor to this linear relationship. Let us suppose that the electronics firm had an exceptionally good year and that it decided to give each of its employees a Christmas bonus of \$500. The equation would now read

$$Y = 500 + 12X.$$

Perhaps, thinking back to your high school days of algebra, you will recognize the above formula as a special case of the general formula for a straight line, that is,

$$Y = a + b_y X, \tag{9.1}$$

in which Y and X represent variables which change from individual to individual, and a and b_y represent constants for a particular set of data. More

specifically, b_y represents the slope of a line relating values of Y to values of X. This is referred to as the regression of Y on X. In the present example, the slope of the line is 12 which means that Y changes by a factor of 12 for each change in X. The letter a represents the value of Y when $X = 0$.

You may also note that the above formula may be regarded as a method for predicting Y from known values of X. When the correlation is 1.00 (as in the present case), the predictions are perfect.

9.2.2 Predicting *X* and *Y* from Data on Two Variables

In behavioral research, however, the correlations we obtain are almost never perfect. Therefore we must find a straight line which best fits our data and we must make predictions from that line. But what do we mean by "best fit?" You will recall that when discussing the mean and the standard deviation, we defined the mean as that point in a distribution that makes the sum of squares of deviations from it minimal (least sum squares). When applying the least sum square method to correlation and regression, the *best fitting straight line* is defined as that line which makes the squared deviations around it minimal. This straight line is referred to as a *regression line.*

We might note at this time that the term prediction, as employed in statistics, does not carry with it any necessary implication of futurity. The term "predict" simply refers to the fact that we are using information about one variable to obtain information about another. Thus, if we know a student's grade point average in college, we may use this information to "predict" his intelligence (which in our more generous moods, we assume preceded his entrance into college).

At this point, we shall introduce two new symbols: X' and Y'. These may be read as "X prime and Y prime," "X and Y predicted," or "estimated X and Y." We use these symbols whenever we employ the regression line or the regression equation to estimate or predict a score on one variable from a known score on another variable.

Returning to the formula for a straight line, we are faced with the problem of determining b and a for a particular set of data so that Y' may be obtained.

The formula for obtaining the slope of the line relating Y to X, which is known as the line of regression of Y on X, is

$$b_y = \frac{\sum xy}{\sum x^2}. \tag{9.2}$$

From formula (9.2) we may derive another useful formula for determining the slope of the line of Y on X.

$$b_y = \frac{\sum xy}{\sum x^2}$$

But

$$\sum xy = r\sqrt{\sum x^2 \cdot \sum y^2}$$

and

$$\sum x^2 = Ns_x^2$$
$$\sum y^2 = Ns_y^2.$$

Thus:

$$b_y = \frac{r\sqrt{N^2 s_x^2 \cdot s_y^2}}{Ns_x^2}$$
$$= r\,\frac{Ns_x s_y}{Ns_x^2}$$
$$= r\,\frac{s_y}{s_x}. \tag{9.3}$$

Similarly, the slope of the regression line of X on Y may be expressed as:

$$b_x = \frac{\sum xy}{\sum y^2} \tag{9.4}$$

or

$$b_x = r\,\frac{s_y}{s_x}. \tag{9.5}$$

It will be noted that all the quantities shown in formulas (9.2) through (9.5) may be readily obtained in the course of calculating Pearson r.

The constant a is given by the formula

$$a = \overline{Y} - b_y\overline{X} \tag{9.6}$$

In the computation of Y', it is unwieldy to obtain each of these values separately and substitute them into the formula for a straight line. However, by algebraically combining formulas (9.3) and (9.6) and relating the result to formula (9.1), we obtain a much more useful formula for Y':

$$Y' = \overline{Y} + r\,\frac{s_y}{s_x}\,(X - \overline{X}).^* \tag{9.7}$$

---

* Since $Y' = a + b_yX$, $a = \overline{Y} - b_y\overline{X}$, and $b_y = r(s_y/s_x)$, then

$$Y' = \overline{Y} - r\left(\frac{s_y}{s_x}\right)\overline{X} + r\left(\frac{s_y}{s_x}\right)X = \overline{Y} + r\left(\frac{s_y}{s_x}\right)(X - \overline{X}).$$

Since there is also a separate regression equation for predicting scores on the X-variable from values of the Y-variable, the formula for X' is

$$X' = \overline{X} + r\frac{s_x}{s_y}(Y - \overline{Y}). \tag{9.8}$$

Concentrating our attention upon the second term on the right of each equation, we can see that the larger the r, the greater the magnitude of the entire term. This term also represents the *predicted deviation from the sample mean resulting from the regression of Y on X or X on Y.* Thus, we may conclude that the greater the correlation, the greater the predicted deviation from the sample mean. In the event of a perfect correlation, the entire predicted deviation is maximal. On the other hand, when $r = 0$, the predicted deviation is also zero. Thus, when $r = 0$, we have $X' = \overline{X}$, and $Y' = \overline{Y}$. All of this is another way of saying that in the absence of a correlation between two variables, our best prediction of any given score on a specified variable is the mean of the distribution of that variable.

9.2.3 Illustrative Regression Problems

Let us solve two sample problems employing the data introduced in Section 9.1.

Problem 1

Mr. Jones, you will recall, scored 62 on the midterm examination. What is our prediction concerning his score on the final examination? The relevant statistics are reproduced below.

X Midterm		Y Final
$\overline{X} = 70$		$\overline{Y} = 75$
$s_x = 4$		$s_y = 8$
	$r = 0.60$	

Employing formula (9.7), we find

$$\begin{aligned} Y' &= 75 + 0.60\,(\tfrac{8}{4})\,(62–70) \\ &= 75 - 9.60 = 65.40. \end{aligned}$$

Problem 2

Mr. Smith, on the same midterm test, scored 76. What is our prediction concerning his score on the final examination? Employing the data appearing in

Problem 1, we obtain the following results:

$$\begin{aligned} Y' &= 75 + 0.60\ (\tfrac{8}{4})\ (76\text{–}70) \\ &= 75 + 7.20 = 82.20. \end{aligned}$$

Had our problem been to "predict" X-scores from known values of Y, the procedures would have been precisely the same as above except that formula (9.8) would have been employed.

A reasonable question at this point is, "Since we know $\overline{Y}$ and s_y in the above problems, we presumably have all the observed data at hand. Therefore, why do we wish to predict Y from X?" It should be pointed out that the purpose of these examples was to acquaint you with the prediction formulas. In actual practice, however, correlational techniques are most commonly employed in making predictions about future samples where Y is unknown.

For example, let us suppose that the admissions officer of a college has constructed an entrance examination which he has administered to all the applicants over a period of years. During this time he has accumulated much information concerning the relationship between entrance scores and subsequent quality point averages in school. He finds that it is now possible to use scores on the entrance examination (X-variable) to predict subsequent quality point averages (Y-variable), and then use this information to establish an entrance policy for future applicants.

Since we have repeatedly stressed the relationship between Pearson r and z-scores, it should be apparent that the prediction formulas may be expressed in terms of z-scores. Mathematically, it can be shown that

$$z_{y'} = rz_x,^* \tag{9.9}$$

where $z_{y'} = Y'$ expressed in terms of a z-score.

Returning to Problem 1, Mr. Jones' score of 62 on the midterm can be expressed as a $z = -2.00$. Thus $z_{y'} = 0.60(-2.00) = -1.20$.

To assure yourself of the comparability of the two prediction formulas, that is, Eqs. (9.7) and (9.9), you should translate the $z_{y'}$-score into a raw score, Y'.

---

* $Y' = \overline{Y} + r(s_y/s_x)(X - \overline{X})$. By transposing terms:

$$\frac{Y' - \overline{Y}}{s_y} = r\,\frac{(X - \overline{X})}{s_x}, \quad \text{but} \quad \frac{X - \overline{X}}{s_x} = z_x \quad \text{and} \quad \frac{Y' - \overline{Y}}{s_y} = z_{y'}.$$

Therefore $z_{y'} = rz_x$.

9.2.4 Constructing Lines of Regression

Let us return to the problem of constructing regression lines for predicting scores on the variables X and Y. As we have already pointed out, the regression line will not pass through all the paired scores. It will, in fact, pass among the paired scores in such a way as to minimize the squared deviations between the regression line (predicted scores) and the obtained scores. Earlier we pointed out that the mean is the point in a distribution which makes the squared deviations around it minimal. In discussing regression, the regression line is analogous to the mean, since, as we shall demonstrate, the sum of deviations of scores around the regression line is zero and the sum squares of these deviations are minimal.

It will be recalled that all the values required to calculate predicted scores are readily found during the course of calculating r, that is, $\overline{X}$, $\overline{Y}$, s_x, s_y. Now to construct our regression line for predicting Y from X, all we need to do is take two extreme values of X, predict Y from each of these values, and then join these two points on the scatter diagram. The line joining these points represents the regression line for predicting Y from X, which is also referred to as the line of regression of Y on X. Similarly, to construct the regression line for predicting X from Y, we take two extreme values of Y, predict X for each of these values, and then join these two points on the scatter diagram. This is precisely what was done in Fig. 9.2 to construct the two regression lines from the data in Table 8.3.

You will note that both regression lines intersect at the means of X and Y. In conceptualizing the relationship between the regression lines and the magnitude of r, it is helpful to think of the regression lines as rotating about the joint means of X and Y. When $r = 1.00$, both regression lines will have identical slopes and will be superimposed upon one another since they pass

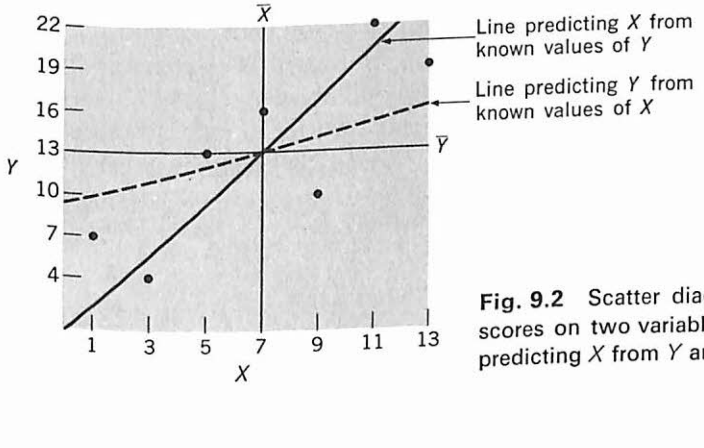


Fig. 9.2 Scatter diagram representing paired scores on two variables and regression lines for predicting X from Y and Y from X.

directly through all the paired scores. However, as r becomes increasingly small, the regression lines rotate away from each other so that in the limiting case when $r = 0$, they are perpendicular to each other. At this point the regression line for predicting X from known values of Y is $\overline{X}$, and the regression line for predicting Y from known values of X is $\overline{Y}$.

9.3 RESIDUAL VARIANCE AND STANDARD ERROR OF ESTIMATE

Figure 9.3 shows a series of scatter diagrams, each reproduced from Fig. 9.2, showing only one regression line: the line for predicting Y from known values of X. Although our present discussion will be directed only to this regression line, all the conclusions we draw will be equally applicable to the line predicting X from known values of Y.

The regression line represents our best basis for predicting Y scores from known values of X. As we can see, not all the obtained scores fall on the regression line. However, if the correlation had been 1.00, all the scores would have fallen right on the regression line. The deviations $(Y - Y')$ in Fig. 9.3 represent our errors in prediction.

You will note the similarity of $Y - Y'$ (the deviation of scores from the regression line) to $Y - \overline{Y}$ (the deviation of scores from the mean). The algebraic sum of these deviations around the regression line is equal to zero.

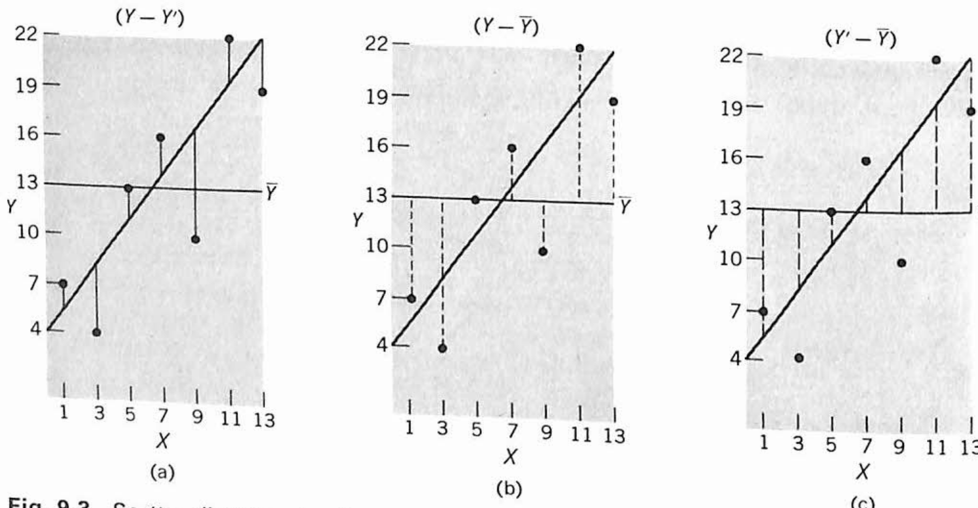


Fig. 9.3 Scatter diagram of paired scores on two variables, regression line for predicting Y-values from known values of X, and the mean of the distribution of Y-scores ($\overline{Y}$); $r = 0.82$ (from data in Table 8.3).

Earlier, we saw that the algebraic sum of the deviations around the mean is also equal to zero. In a sense, then, the regression line is a sort of "floating mean:" one that takes on different values depending on the values of X that are employed in prediction.

You will also recall that in calculating the variance, s^2, we squared the deviations from the mean, summed, and divided by N. Finally, the square root of the variance provided our standard deviation. Now, if we were to square and sum the deviations of the scores from the regression line, $\sum(Y - Y')^2$, we would have a basis for calculating another variance and standard deviation. The variance around the regression line is known as the *residual variance* and is defined as follows:

$$s^2_{\text{est}\,y} = \frac{\sum(Y - Y')^2}{n}. \tag{9.10}$$

When predictions are made from Y to X, the residual variance of X is

$$s^2_{\text{est}\,x} = \frac{\sum(X - X')^2}{n}. \tag{9.11}$$

The standard deviation around the regression line (referred to as the *standard error of estimate*) is, of course, the square root of the residual variance. Thus

$$s_{\text{est}\,y} = \sqrt{\frac{\sum(Y - Y')^2}{n}}. \tag{9.12}$$

When predictions are made from Y to X, the standard error of estimate of X is

$$s_{\text{est}\,x} = \sqrt{\frac{\sum(X - X')^2}{n}}. \tag{9.13}$$

You may be justifiably aghast at the amount of computation that is implied in the above formulas for calculating the standard error of estimate. However, as has been our practice throughout this text, we have shown the basic formulas so that you may have a conceptual grasp of the meaning of the standard error of estimate. It is, as we have seen, the standard deviation of scores around the regression line rather than around the mean of the distribution.

Fortunately, as in all previous illustrations in the text, there is a simplified method for calculating $s_{\text{est}\,y}$ and $s_{\text{est}\,x}$:

$$s_{\text{est}\,y} = s_y\sqrt{1 - r^2},^* \tag{9.14}$$

and

$$s_{\text{est}\,x} = s_x\sqrt{1 - r^2}.^* \tag{9.15}$$

* The values of $\sqrt{1 - r^2}$ may be obtained directly from Table H. Thus for $r = 0.82$, we have $\sqrt{1 - r^2} = (0.5724)$.

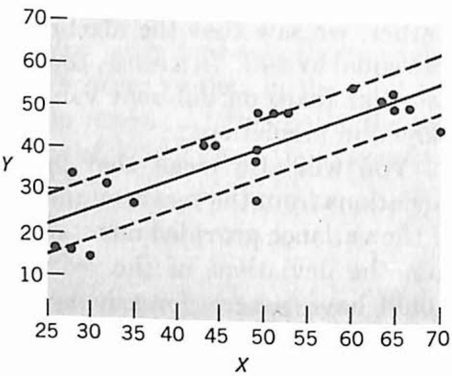


Fig. 9.4 Line of regression for predicting Y from X with parallel lines $1s_{\text{est}\,y}$ above and below the regression line (from data in Problem 1, Chapter 8). Circles indicate individuals' scores on X and Y.

You will note that when $r = \pm 1.00$, $s_{\text{est}\,y} = 0$, which means there are no deviations from the regression line, and therefore no errors in prediction. On the other hand, when $r = 0$, the errors of prediction are maximal for that given distribution, that is, $s_{\text{est}\,y} = s_y$.

With the data in Problem 1, Chapter 8, the following statistics were calculated:

College entrance exam X	Verbal comprehension exam Y
$\overline{X} = 47.65$	$\overline{Y} = 39.15$
$s_x = 13.82$	$s_y = 12.35$

$$r = 0.85.$$

Thus

$$\begin{aligned} s_{\text{est}\,y} &= 12.35\sqrt{1 - 0.84^2} \\ &= 12.35(0.5268) = 6.51. \end{aligned}$$

As already indicated, the standard error of estimate has properties that are similar to those of the standard deviation. For example, if we were to construct lines parallel to the regression line for predicting Y from X at distances $1s_{\text{est}\,y}$, $2s_{\text{est}\,y}$, and $3s_{\text{est}\,y}$, we would find that approximately 68% of the cases fall between $\pm 1s_{\text{est}\,y}$, 95% between $\pm 2s_{\text{est}\,y}$, and 99% between $\pm 3s_{\text{est}\,y}$. These relationships between standard error of estimate and percentage of area will be more closely approximated when the variability within the columns and the rows is homogeneous. This condition is known as *homoscedasticity*.

Using the above data, we have drawn two lines parallel to the regression line for predicting Y from known values of X in Fig. 9.4. These lines are both 1 standard error of estimate (± 6.51) from the regression line. Each circle represents an individual's scores on the X- and the Y-variables. It can be seen that 13 of the 20 scores, or 65% of the cases, fall between $\pm 1s_{\text{est}\,y}$. This figure is in fairly good agreement with the expected percentage of 68. With a larger n, the approximation would be better.

9.4 EXPLAINED AND UNEXPLAINED VARIATION*

If we look again at Fig. 9.3, we can see that there are three separate sum squares that can be calculated from the data. These are:

1. Variation of scores around the sample mean (Fig. 9.3b). This variation is given by $(Y - \overline{Y})^2$ and is, of course, basic to the determination of the variance and the standard deviation of the sample.

2. Variation of scores around the regression line (or predicted scores) (Fig. 9.3a). This variation is given by $(Y - Y')^2$ and is referred to as *unexplained variation*. The reason for this choice of terminology should be clear. If the correlation between two variables is ± 1.00, all the scores fall on the regression line. Consequently, we have, in effect, explained *all* the variation in Y in terms of the variation in X and, conversely, all the variation of X in terms of the variation in Y. In other words, in the event of a perfect relationship, there is no unexplained variation. However, when the correlation is less than perfect, many of the scores will not fall right on the regression line, as we have seen. The deviations of these scores from the regression line represent variation which is not accounted for in terms of the correlation between two variables. Hence, the term "unexplained variation" is employed.

3. Variation of predicted scores about the mean of the distribution (Fig. 9.3c). This variation is given by $(Y' - \overline{Y})^2$ and is referred to as *explained variation*. The reason for this terminology should be clear from our discussion in the preceding paragraph and our prior reference to predicted deviation (Section 9.3). You will recall our previous observation that the greater the correlation, the greater the predicted deviation from the sample mean. It follows further that the greater the predicted deviation, the greater the explained variation. When the predicted deviation is maximum, the correlation is perfect, and the explained variation is 100%.

It can be shown mathematically that the total sum squares consists of two components which may be added together. These two components represent explained variation and unexplained variation respectively. Thus

$$\sum(Y - \overline{Y})^2 = \sum(Y - Y')^2 + \sum(Y' - \overline{Y})^2. \tag{9.16}$$

Total variation = unexplained variation + explained variation.

Now, when $r = 0.00$, then $\sum(Y' - \overline{Y})^2 = 0.00$. (Why? See Section 9.2). Consequently, the total variation is equal to the unexplained variation. Stated another way, when $r = 0$, all the variation is unexplained. On the other hand, when $r = 1.00$, then $\sum(Y - Y')^2 = 0.00$, since all the scores are on the re-

* Although analysis of variance is not covered until Chapter 15, much of the material in this section will serve as an introduction to some of the basic concepts of analysis of variance.

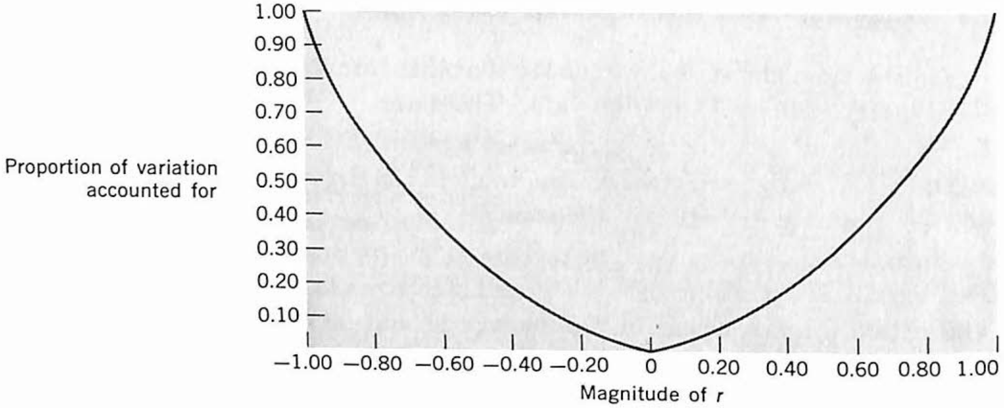


Fig. 9.5 The proportion of the variation on one variable accounted for in terms of variations of a correlated variable at varying values of r.

gression line. Under these circumstances, total variation is the same as explained variation. In other words, all the variation is explained when $r = 1.00$.

The ratio of the explained variation to the total variation is referred to as the *coefficient of determination* and is symbolized by r^2. The formula for the coefficient of determination is

$$r^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{\sum(Y' - \overline{Y})^2}{\sum(Y - \overline{Y})^2}. * \tag{9.17}$$

It can be seen that the coefficient of determination indicates the proportion of total variation which is explained in terms of the magnitude of the correlation coefficient. When $r = 0$, the coefficient of determination r^2, equals 0. When $r = 0.5$, the coefficient of determination is 0.25. In other words, 25% of the total variation is accounted for. Finally, when $r = 1.00$, then $r^2 = 1.00$ and all variation is accounted for.

Figure 9.5 depicts graphically the proportion of variation in one variable that is accounted for by the variation in another variable when r takes on different values.

You have undoubtedly noted that the square root of the coefficient of determination provides another definition of r. Thus

$$r = \pm\sqrt{\frac{\text{explained variation}}{\text{total variation}}} = \pm\sqrt{\frac{\sum(Y' - \overline{Y})^2}{\sum(Y - \overline{Y})^2}}. \tag{9.18}$$

* Table H in Appendix III presents a number of functions of r for various values of r, including such useful functions as r^2, $1 - r^2$, and $\sqrt{1 - r^2}$. You should familiarize yourself with this table.

Since r^2 represents the proportion of variation accounted for, $(1 - r^2)$ represents the proportion of variation which is not explained in terms of the correlation between X and Y. This concept is known as the *coefficient of nondetermination* and is symbolized by k^2. Thus k^2 represents the proportion of variation in Y which must be explained by variables other than X.

Summarizing the relationship between k^2 and r^2:

$$k^2 = 1 - r^2 \tag{9.19}$$

or

$$k^2 + r^2 = 1. \tag{9.20}$$

9.5 CORRELATION AND CAUSATION

You have seen that when two variables are related it is possible to predict one from your knowledge of the other. This relationship between correlation and prediction often leads to a serious error in reasoning, i.e., the relationship between two variables frequently carries with it the implication that one has caused the other. This is especially true when there is a temporal relationship between the variables in question, i.e., when one precedes the other in time. What is often overlooked is the fact that the variables may not be causally connected in any way, but that they may vary together by virtue of a common link with a third variable. Thus, if you are a bird watcher, you may note that as the number of birds increases in the spring, the grass becomes progressively greener. However, recognizing that the extended number of hours and the increasing warmth of the sun is a third factor influencing both of these variables, you are not likely to conclude that the birds cause the grass to turn green or vice versa. However, there are many occasions, particularly in the behavioral sciences, when it is not so easy to identify the third factor.

Suppose that you have demonstrated that there is a high positive correlation between the number of hours students spend studying for an exam and their subsequent grades on that exam. You may be tempted to conclude that the number of hours of study causes grades to vary. This seems to be a perfectly reasonable conclusion, and is probably in close agreement with what your parents and instructors have been telling you for years. Let us look closer at the implications of a causal relationship. On the assumption that a greater number of hours of study causes grades to increase, we would be led to expect that *any* student who devotes more time to study is guaranteed a high grade and that one who spends less time with his books is going to receive a low grade. This is not necessarily the case. We have overlooked the fact that it might be that the better student (by virtue of higher intelligence, stronger

motivation, better study habits, etc.) who devotes more time to study, performs better simply because he has a greater capacity to do so.

What we are saying is that correlational studies simply do not permit inferences of causation. Correlation is a necessary but not a sufficient condition to establish a causal relationship between two variables. In short, to establish a causal relationship it is necessary to conduct an experiment in which an independent variable is manipulated by the experimenter, and the effects of these manipulations are reflected in the dependent, or criterion, variable. A correlational study lacks the requirement of independent manipulation.

Huff's book* includes an excellent chapter devoted to the confusion of correlation with causation. He refers to faulty causal inferences from correlational data as the *post hoc* fallacy. The following excerpt illustrates a common example of the *post hoc* fallacy.

> Reams of pages of figures have been collected to show the value in dollars of a college education, and stacks of pamphlets have been published to bring these figures —and conclusions more or less based on them— to the attention of potential students. I am not quarreling with the intention. I am in favor of education myself, particularly if it includes a course in elementary statistics. Now these figures have pretty conclusively demonstrated that people who have gone to college make more money than people who have not. The exceptions are numerous, of course, but the tendency is strong and clear.
>
> The only thing wrong is that along with the figures and facts goes a totally unwarranted conclusion. This is the *post hoc* fallacy at its best. It says that these figures show that if *you* (your son, your daughter) attend college you will probably earn more money than if you decide to spend the next four years in some other manner. This unwarranted conclusion has for its basis the equally unwarranted assumption that since college trained folks make more money, they make it because they went to college. Actually we don't know but that these are the people who would have made more money even if they had not gone to college. There are a couple of things that indicate rather strongly that this is so. Colleges get a disproportionate number of two groups of kids—the bright and the rich. The bright might show good earning power without college knowledge. And as for the rich ones . . . well money breeds money in several obvious ways. Few sons of rich men are found in low-income brackets whether they go to college or not.

CHAPTER SUMMARY

Let us briefly review what we have learned in this chapter. We have seen that it is possible to "fit" two straight lines to a bivariate distribution of scores, one for predicting Y-scores from known X-values and one for predicting X-scores from known Y-values.

We saw that, when the correlation is perfect, all the scores fall upon the regression line. There is, therefore, no error in prediction. The lower the rela-

* *Op. cit.*

tionship, the greater the dispersion of scores around the regression line, and the greater the errors of prediction. Finally, when $r = 0$, the mean of the sample provides our "best" predictor for a given variable.

The regression line was shown to be analogous to the mean: the summed deviations around it are zero and the sum squares are minimal. The standard error of estimate was shown to be analogous to the standard deviation.

We saw that three separate sum squares, reflecting variability, may be calculated from correlational data.

1. Variation about the mean of the distribution for each variable. This variation is referred to as the *total sum squares.*
2. Variation of each score about the regression line. This variation is known as *unexplained variation.*
3. Variation of each predicted score about the mean of the distribution for each variable. This variation is known as *explained variation.*

We saw that the sum of the explained variation and the unexplained variation is equal to the total variation.

Finally, we saw that the ratio of the explained variation to the total variation provides us with the proportion of the total variation which is explained. The term applied to this proportion is *coefficient of determination.* In addition, the converse concept of coefficient of nondetermination was discussed.

Terms to Remember:

Line of "best fit"
Regression line
Prediction
Regression equation
Homoscedasticity
Standard error of estimate
"Floating mean"
Residual variance
Unexplained variation
Explained variation
Coefficient of determination
Coefficient of nondetermination
"Post hoc" fallacy

EXERCISES

1. Find the equation of the regression line for the following data.

X	1	2	3	4	5
Y	5	3	4	2	1

2. In a study concerned with the relationship between two variables, X and Y, the following was obtained,

$$\overline{X} = 119 \qquad \overline{Y} = 1.30$$
$$s_x = 10 \qquad s_y = 0.55$$
$$r = 0.70$$
$$n = 100$$

a) Sally B. obtained a score of 130 on the X-variable. Predict her score on the Y-variable.
b) A score of 1.28 on the Y-variable was predicted for Bill B. What was his score on the X-variable?
c) Determine the standard error of estimate of Y.

3. A study was undertaken to find the relationship between "emotional stability" and performance in college. The following results were obtained.

Emotional stability	College average
$\overline{X} = 49$	$\overline{Y} = 1.35$
$s_x = 12$	$s_y = 0.50$

$$r = 0.36$$
$$n = 60$$

a) Norma obtained a score of 65 on the X-variable. What is our prediction of her score on the Y-variable?
b) Determine the standard error of estimate of X and Y.
c) What proportion of total variation is accounted for by explained variation?

4. Assume that $\overline{X} = 30$, $s_x = 5$; $\overline{Y} = 45$, $s_y = 8$. Draw a separate graph for each pair of regression lines for the following values of r.

a) 0.00 b) 0.20 c) 0.40
d) 0.60 e) 0.80 f) 1.00

Generalize: What is the relationship between the size of r and the angle formed by the regression lines? If the values of r given above, (b) through (f), were all negative, what is the relationship?

5. Given: The standard deviation of scores on a standardized vocabulary test is 15. The correlation of this test with I.Q. is 0.80. What would you expect the standard deviation on the vocabulary test to be for a large group of students with the same I.Q.? Explain your answer.

6. A student obtains a score on test X which is 1.5 standard deviations above the mean. What standard score would you predict for him on test Y if r equals
a) 0.00, b) 0.40, c) 0.80, d) 1.00, e) —0.50, f) —0.80.

7. A personnel manager has made a study of employees involved in one aspect of a manufacturing process. He finds that after they have been on the job for a year, he is able to obtain a performance measure which accurately reflects their proficiency. He designs a selection test aimed at predicting their eventual proficiency, and obtains a correlation of 0.65 with the proficiency measure (Y). The mean of the test is 50, $s_x = 6$; $\overline{Y} = 100$, $s_y = 10$. Answer the following questions based on these facts:

a) Herman J. obtained a score of 40 on the selection test. What is his predicted proficiency score?
b) How likely is it that he will score as high as 110 on the proficiency scale?
c) A score of 80 on the Y-variable is considered satisfactory for the job; below 80 is unsatisfactory. If the X-test is to be used as a selection device, which score should be used as a cutoff point? (*Hint:* Find the value of X which leads to a prediction of 80 on Y. Be sure to employ the appropriate prediction formula.)

d) Sonya J. obtained a score of 30 on X. How likely is it that she will achieve an acceptable score on Y?

e) Leon M. obtained a score of 60 on X. How likely is it that he will *fail* to achieve an acceptable score on Y?

f) For a person to have prospects for a supervisory position, a score of 120 or higher on Y is deemed essential. What value of X should be employed for the selection of potential supervisory personnel?

g) If 1000 persons achieve a score on X which predicts a $Y = 120$, approximately how many of them will obtain Y scores below 120? Above 130? Below 110? Above 110?

8. An owner of a mail-order house advertises that all orders are shipped within 24 hr of receipt. Since personnel in the shipping department are hired on a day-to-day basis, it is important for him to be able to predict the number of orders contained in each batch of daily mail so that he can hire sufficient personnel for the following day. He hit on the idea of weighing each day's mail and correlating the weight with the actual number of orders. Over a successive 30-day period, he obtained the following results:

Weight in pounds	Number of orders	Weight in pounds	Number of orders
20	5400	26	5400
15	4200	21	5000
23	5800	24	5400
17	5000	16	4300
12	3500	34	6700
35	6400	28	6100
29	6000	15	3600
21	5200	11	3200
10	4000	18	5300
13	3800	27	5800
25	5700	30	5900
14	4000	22	5500
18	4800	20	5200
30	6200	24	5000
33	6600	13	3700

a) Find the correlation between the weight of mail and the number of orders. (*Hint:* In calculating the correlation, consider dropping the final two digits on the Y-variable).

b) If 10 persons are required to handle 1000 orders per day, how many should be hired to handle 22 pounds of mail? 15? 30? 38? (*Note:* Assume that employees are hired only in groups of 10).

9. Peruse the magazine section of your Sunday newspaper, monthly magazines, television, and radio advertisements for examples of the *post hoc* fallacy.

10. Referring back to Problem 1, Chapter 8:
   a) Estelle obtained a score of 40 on her college entrance examination. Predict her score on the verbal comprehension test.
   b) How likely is it that she will score at least 40 on the verbal comprehension test?
   c) Howard obtained a score of 40 on the verbal comprehension test. Predict his score on the college entrance examination.
   d) How likely is it that he will score at least 40 on the college entrance examination?
   e) REC University finds that students who score at least 45 on the verbal comprehension test are most successful. What score on the college entrance examination should be used as the cutoff point for selection?
   f) Harris obtained a score of 55 on the college entrance examination. Would he be selected by REC University? What are his chances of achieving an acceptable score on the verbal comprehension test?
   g) Rona obtained a score of 60 on the college entrance examination. Would she be accepted by REC University? How likely is it that he will *not* achieve an acceptable score on the verbal comprehension test?
11. On the basis of the obtained data (below) an experimenter asserts that the older a child is the fewer irrelevant responses he makes in an experimental situation.
   a) Determine whether this conclusion is valid.
   b) Mindy, age 13, enters the experimental situation:
      i) What is the most probable number of irrelevant responses the experimenter would perdict for Mindy?
      ii) What is the likelihood that she will make no irrelevant responses?

Age	Number irrelevant responses	Age	Number irrelevant responses
2	11	7	12
3	12	9	8
4	10	9	7
4	13	10	3
5	11	11	6
5	9	11	5
6	10	12	5
7	7		

12. Why do we have two regression lines?

Review of Section I

DESCRIPTIVE STATISTICS

In the preceding 9 chapters you have seen that there are many different ways to describe data. The application of the various descriptive techniques on the following two sets of data will give you an opportunity to see how these techniques can be integrated to describe data.

Review Problem 1

A. Draw a *graph* to illustrate the relationship between the final standing and the number of home runs hit.

B. Calculate:
   1) *Mean* number of times the American League teams were shut out.
   2) *Median* number of home runs hit in the American League.
   3) *Standard deviation* of the distribution of shut-outs.

C. Determine the relationship between
   1) the number of home runs and the final standing.
   2) the number of home runs and the number of times shut-out.

Final standings	Home runs	Times shut out
Minnesota	150	3
Chicago	125	9
Baltimore	125	12
Detroit	162	11
Cleveland	156	15
New York	149	16
California	92	16
Washington	136	14
Boston	165	11
Kansas City	110	10

Review Problem 2*

A. Draw a graph to illustrate the changes from 1946–1963 in average daily cost per hospital patient.
B. Draw a frequency distribution of the average length of hospital stay from the data presented below. Describe the distribution.
C. For the 18-year period from 1946–1963, calculate the mean and standard deviation of the average length of hospital stay.
D. Determine the relationship between the average cost per patient per hospital day and the average length of hospital stay.

	Average cost per patient per hospital day	Average length of hospital stay (in days)		Average cost per patient per hospital day	Average length of hospital stay (in days)
1946	$ 9.39	9.1	1955	$23.12	7.8
1947	11.09	8.0	1956	24.15	7.7
1948	13.09	8.7	1957	26.02	7.6
1949	14.33	8.3	1958	28.27	7.6
1950	15.62	8.1	1959	30.19	7.8
1951	16.77	8.3	1960	32.23	7.6
1952	18.35	8.1	1961	34.98	7.6
1953	19.95	7.9	1962	36.83	7.6
1954	21.76	7.8	1963	38.91	7.7

1) A person who spends one week (7 days) in a hospital can expect his stay to average how much daily?
2) On the basis of these calculations, patient X concludes that the longer he stays in the hospital the less it will cost him per day. He decides that if he can prolong his hospital stay for 10 days he can expect an average daily cost of ______? What is the fallacy in his conclusion? [*Hint:* Criticize this inference in the light of what you know about correlation and causation.]

---

* The data for this problem were adapted from *Reader's Digest Almanac*, p. 489. New York: Reader's Digest Association, 1966, with permission.

Section II
Inferential Statistics

A. PARAMETRIC TESTS OF SIGNIFICANCE

Probability 10

10.1 AN INTRODUCTION TO PROBABILITY THEORY

In the past few chapters, we have been primarily concerned with the exposition of techniques employed by statisticians and scientists to describe and present data in the most economical and meaningful form. However, we pointed out in Chapter 1, that the interests of scientists go beyond the mere description of data. Fundamental to the strategy of science is the formulation of general statements about populations or the effects of experimental conditions on criterion variables. Thus, as we have already pointed out, the scientist is not usually satisfied to report merely that the arithmetic mean of the drug group tested on variable X is higher or lower than the mean of the placebo group tested on this variable. He also wants to make general statements such as: "The difference between the two groups is of such magnitude that we cannot reasonably ascribe it to chance variation. We may therefore conclude that the drug had an effect on the variable studied. More specifically, this effect was . . . etc., etc."

The problem of chance variation is an important one. We all know that the variability of our data in the behavioral sciences engenders the risk of drawing an incorrect conclusion. Take a look at the following example: From casual observations, Experimenter A hypothesizes that first grade girls have higher I.Q. scores than first grade boys. He administers an I.Q. test to four boys and four girls in a first grade class. He finds that the mean of the girls is higher: 110 to 103. Is Experimenter A justified in concluding that his hypothesis has been confirmed? The answer is obviously in the negative. But why? After all, there is a difference between the sample means, isn't there? Intuitively, we might argue that the variability of intelligence among first graders is so great, and the N in the study so small, that *some* differences in the means is inevitable as a result of our selection procedures. The critical questions which must be answered by inferential statistics then become: "Is the apparent difference in intelligence among first graders reliable? That is, will it appear regularly in

repetitions of the study? Or is the difference the result of unsystematic factors which will vary from study to study, and thereby produce sets of differences without consistency?" A prime function of inferential statistics is to provide rigorous and logically sound procedures for answering these questions. As we shall see in this chapter and the next, probability theory provides the logical basis for deciding among all the various alternative interpretations of research data.

Probability theory is not as unfamiliar as many would think. Indeed, in everyday life we are constantly called upon to make probability judgments although we may not recognize them as such.

For example, let us suppose that, for various reasons, you are unprepared for today's class. You seriously consider not attending class. What are the factors that will influence your decision? Obviously, one consideration would be the likelihood that the instructor will detect your lack of preparation. If the risk is high, you decide not to attend class; if low, then you will attend.

Let us look at this example in slightly different terms. There are two alternative possibilities

event A: Your lack of preparation *will* be detected.

event B: Your lack of preparation *will not* be detected.

There is uncertainty in this situation because more than one alternative is possible. Your decision whether or not to attend class will depend upon your degree of assurance associated with each of these alternatives. Thus, if you are fairly certain that the first alternative will prevail, you will decide not to attend class.

Suppose that your instructor frequently calls upon students to participate in class discussion. In fact, you have noted that most of the students are called upon in any given class session. This is an example of a situation in which a high degree of assurance is associated with the first alternative. Stated another way, the probability of event A is higher than the probability of event B. Thus, you decide not to attend class.

Although you have not used any formal probability laws in this example, you have actually made a judgment based upon an intuitive use of probability.

You may be aware of the fact that many of the questions raised in the exercises began with, "What is the likelihood that . . .?" These questions were in preparation for the formal discussion of probability occurring in the present and subsequent chapters. However, before discussing the elements of probability theory, it is desirable to understand one of the most important concepts in inferential statistics, that of *randomness*.

10.2 THE CONCEPT OF RANDOMNESS

A series of events are said to be random if one event has no predictable effect on the next. We can most readily grasp randomness in terms of games of chance, assuming they are played honestly. Knowledge of the results of one toss of a coin, one throw of a die, one outcome on the roulette wheel, or one selection of a card from a well-shuffled deck (assuming replacement of the card after each selection) will not aid us one iota in our predictions of future outcomes. This characteristic of random events is known as *independence*. Only if independence is achieved, can events be said to be truly random. The second important characteristic of randomness is that when the sample is taken from a population, each member of the population should have an *equally likely* chance to be selected. Thus, if our selections favor certain events or certain collections of events, we cannot justifiably claim randomness. Such sampling procedures are referred to as being *biased*. In the naturalistic type of study, alluded to earlier, in which our purpose is to describe certain characteristics of a population, the problem of bias is an ever-present danger. When we are interested in learning the characteristics of the general population on a given variable, we dare not select our sample from automobile registration lists or "at random" on a street corner in New York City. The dangers of generalizing to the general population from such biased samples should be obvious to you. Unless the condition of randomness is met, we may never know to what population we should generalize our results. Furthermore, with nonrandom samples, we find that many of the rules of probability do not hold.

It is beyond the scope of this text to delve deeply into sampling procedures since this topic is a full course by itself. However, let us look at an illustration of the procedures by which we may achieve randomness in assigning subjects to experimental conditions.

Let us suppose that you are interested in comparing three different methods of teaching reading readiness to pre-school children. There are 87 subjects who are to be divided into three equal groups. The assignment of these subjects must be made in a random manner.

One method to achieve randomness would be to place each subject's name on a slip of paper. We shuffle these slips and then place them into three equal piles.

An alternative method would be to use the Table of Random Digits (Table R in Appendix III). Since the digits in this table have already been randomized, the effect of shuffling has been achieved. We assign each of the 87 students a numeral from 1 to 87. We may start with any row or column of digits in Table R.

After we have selected 29 numerals that correspond to 29 different subjects, we have formed our first group. We continue until we have three groups each consisting of 29 different subjects. For example, if we start with the fifteenth

row and then choose consecutive pairs of digits, we obtain the following subjects: 65, 48, 11, 76, 74, 17, etc. If any numeral over 87 or any repeated numeral appears, we disregard it.

The reason for the paramount importance of random procedures will become clear in this chapter and the next. Fundamentally, it is based on a fascinating fact of inferential statistics; i.e., *each event may not be predictable when taken alone, but collections of random events can take on predictable forms.* The binomial distribution, which we shall discuss at greater length in Section 11.3, illustrates this fact. If we were to take, say, 20 unbiased coins and toss them into the air, we could not predict accurately the proportion that would land "heads." However, if we were to toss these 20 coins a large number of trials, record the number turning up heads on each trial, and construct a frequency distribution of outcomes in which the horizontal axis varies between no heads and all heads, the plot would take on a characteristic and predictable form known as the *binomial* or *Bernoulli distribution*. By employing the Bernoulli model, we would be able to predict with considerable accuracy, over a large number of trials, the percentage of the time various outcomes will occur. The same is true with respect to the normal curve model. In the absence of any specific information, we might not be able to predict a person's status with respect to a given trait (intelligence, height, weight, etc.). However, as we already know, frequency distributions of scores on these traits commonly take the form of the normal curve. Thus, we may predict the proportion of individuals scoring between specified score limits.

What is perhaps of more importance, from the point of view of inferential statistics, is the fact that distributions of sample statistics ($\overline{X}$, s, median, etc.) based on random sampling from a population, also take on highly predictable forms. Chapter 11 deals with the concept of *sampling distributions*, which are theoretical probability distributions of a statistic which would result from drawing all possible samples of a given size from some population.

With this brief introduction to the concept of randomness, we are prepared to look at probability theory.

10.3 APPROACHES TO PROBABILITY

Probability may be regarded as a theory that is concerned with the possible outcomes of experiments. The experiments must be potentially repetitive; i.e., we must be able to repeat them under similar conditions. It must be possible to enumerate every outcome that can occur, and we must be able to state the expected relative frequencies of these outcomes.

It is the method of assigning relative frequencies to each of the possible outcomes that distinguishes the classical from the empirical approach to probability theory.

10.3.1 Classical Approach to Probability

The theory of probability has always been closely associated with games of chance. For example, suppose that we want to know the probability that a coin will turn up heads. Since there are only two possible outcomes (heads or tails) we assume an ideal situation in which we expect that each outcome is equally likely to occur. Thus, the probability that heads, $p(H)$, will occur is $\frac{1}{2}$. This kind of reasoning has lead to the following classical definition of probability:

$$p(A) = \frac{\text{no. outcomes favoring event } A}{\text{total no. of events (those favoring } A + \text{those not favoring } A)}. \tag{10.1}$$

It should be noted that probability is defined as a proportion (p). The most important point in the classical definition of probability is the assumption of an *ideal* situation in which the structure of the population is known; i.e., the total number of possible outcomes (N) is known. The expected relative frequency of each of these outcomes is arrived at by deductive reasoning. Thus the probability of an event is interpreted as a theoretical or an idealized relative frequency of the event. In the above example, the total number of possible outcomes was 2 (heads or tails), and the relative frequency of each outcome was assumed to have an equal likelihood of occurrence.* Thus, $p(H) = \frac{1}{2}$ and $p(T) = \frac{1}{2}$.

10.3.2 Empirical Approach to Probability

Although it is usually easy to assign expected relative frequencies to the possible outcomes of games of chance, we cannot do this for most real-life experiments. In actual situations, expected relative frequencies are assigned on the basis of empirical findings. Thus we may not know the exact proportion of students in a university who have blue eyes, but we may study a random sample of students and estimate the proportion who will have blue eyes. Once we have arrived at an estimate, we may employ classical probability theory to answer questions such as: What is the probability that in a sample of 10 students, drawn at random from the student body, three or more will be blue-eyed? Or, what is the probability that student John, drawn at random from the student body, will have blue eyes?

* Expected relative frequencies need not be assumed to be equal. In Chapter 17 we will deal with situations in which the expected relative frequencies are not assumed to be equal.

If, in a random sample of 100 students, we found that 30 had blue eyes, we could estimate that the proportion of blue-eyed students in the university was 0.30 by employing formula (10.1):

$$p \text{ (blue-eyed)} = \tfrac{30}{100} = 0.30.$$

Thus the probability is 0.30 that student John will have blue eyes. *Note:* this represents an *empirical* probability; i.e., the expected relative frequency was assigned on the basis of empirical findings.

Although we employ an idealized model in our forthcoming discussion about the properties of probability, we can apply the same principles to diverse practical problems.

10.4 FORMAL PROPERTIES OF PROBABILITIES

10.4.1 Probabilities Vary Between 0 and 1.00

From the classical definition of probability, p is always a number between 0 and 1 inclusively. If an event is certain to occur, its probability is 1; if it is certain not to occur, its probability is 0. For example, the probability of drawing the ace of spades from an ordinary deck of 52 playing cards is $\frac{1}{52}$. The probability of drawing a red ace of spaces is zero since there are no events favoring this result. If all events favor a result (for example, drawing a card with *some* marking on it) $p = 1$. Thus for any given event, say A,

$$0 \leq p(A) \leq 1.00,$$

in which the symbol $\leq$ means "less than or equal to."*

10.4.2 Expressing Probability

In addition to expressing probability as a proportion, several other ways are often employed. It is sometimes convenient to express probability as a *percentage* or as the *number of chances in* 100.

To illustrate: If the probability of an event is 0.05, we expect this event to occur 5% of the time, or the chances that this event will occur are 5 in 100. This same probability may be expressed by saying that the odds are 95 to 5 *against* the event occurring, or 19 to 1 against it.

* The symbol $\geq$ means "greater than or equal to."

Note that when expressing probability as the *odds against* the occurrence of an event, we use the following formula:

Odds against event A =
(total no. of outcomes − no. favoring event A) *to* no. favoring event A. (10.2)

Thus, if $p(A) = 0.01$, the odds against the occurrence of event A are 99 to 1.

10.4.3 The Addition Rule

Occasionally we want to determine the probability of one of several different events. For example, let us suppose that you draw one card from a well-shuffled deck of 52 playing cards. What is the probability that this card is either a king or a black picture card?* The total number of events is 52 (i.e., $N = 52$). There are 4 kings and 6 black picture cards. This would seem to add up to 10.

We might then suspect that the probability of obtaining either a king or a black picture card is $\frac{10}{52}$. Note, however, that in arriving at this total we have counted some cards twice: the black kings have been counted in *both* the king and the black picture card totals. Obviously, we should count them in only one of the totals. Thus, if we include the black kings in the black picture card totals, our revised figures become: 2 red kings plus 6 black picture cards resulting in a total of 8. Our probability then becomes

$$p = \tfrac{8}{52} = 0.15.$$

Note that another approach would be to add the number of kings and the number of black picture cards ($4 + 6 = 10$), and then subtract the number that *share* both characteristics (2), since we previously added this twice. Thus our probability becomes

$$p = \frac{4 + 6 - 2}{52} = \frac{8}{52} = 0.15.$$

This leads us to the general formulation of the addition rule.

If A and B are events, the probability of obtaining either of them is equal to the probability of A plus the probability of B minus the probability of their joint occurrence.

* The word *or* is used in the mathematical sense of the word and includes the possibility that both events will occur. Thus the word *or* is used in the sense of *and/or*.

In symbolic form, this reads

$$p(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B). \tag{10.3}$$

Applied to the above example,

$$p(A \text{ or } B) = \tfrac{4}{52} + \tfrac{6}{52} - \tfrac{2}{52} = \tfrac{8}{52} = 0.15.$$

Note that if the events A and B are *mutually exclusive* (i.e., if both events *cannot* occur simultaneously), the last term disappears. Thus the addition rule with *mutually exclusive* events becomes

$$p(A \text{ or } B) = p(A) + p(B). \tag{10.4}$$

To illustrate: What is the probability of drawing a spade or a club from a well-shuffled deck? Since a single card cannot be both a spade and a club, these two events are mutually exclusive. Thus employing formula (10.4), we have

$$p(A \text{ or } B) = \tfrac{13}{52} + \tfrac{13}{52} = \tfrac{26}{52} = 0.50.$$

In Chapters 11 and 17, we shall be dealing with problems based on *dichotomous, Yes–No*, or *two-category* populations in which the events in question are not only mutually exclusive, but are *exhaustive.* For example, in the problem, "On one draw from a well-shuffled deck of playing cards, what is the probability of drawing either a red card or a black card?" it is not only impossible to obtain both colors simultaneously (i.e., they are mutually exclusive) but there is no possible outcome other than a black or a red card. In other words, red and black exhaust all the possible outcomes. In the event of *mutually exclusive and exhaustive events,* we arrive at the very useful formulation

$$p(A) + p(B) = 1.00. \tag{10.5}$$

In treating dichotomous populations, it is common practice to employ the two symbols P and Q to represent, respectively, the probability of the occurrence of an event and the probability of the nonoccurrence of an event. Thus, if we are flipping a single coin, we can let P represent the probability of occurrence of a head and Q the probability of the nonoccurrence of a head (i.e., the occurrence of a tail). These considerations lead to three useful formulations:

$$P + Q = 1.00, \tag{10.6}$$

$$P = 1.00 - Q, \tag{10.7}$$

$$Q = 1.00 - P, \tag{10.8}$$

when the events are *mutually exclusive and exhaustive.*

10.4.4 The Multiplication Rule

Sometimes, we are faced with the problem of ascertaining the probability of the *joint occurrence* of two or more events. For example, we might wish to determine the probability of selecting a person from the general population, who has an I.Q. score of 130 or more *and* who has an income in excess of $10,000 a year. It will be recognized that this problem differs from the one in which the addition rule is applicable. Instead of being interested in determining the probability of obtaining at least one of the two outcomes, we want to know the probability of obtaining both outcomes simultaneously. To obtain this probability value, we must apply the multiplication rule:

> *Given two events A and B, the probability of obtaining both A and B jointly is the product of the probability of obtaining one of these events times the conditional probability of obtaining one event, given that the other event has occurred.*

Stated symbolically,

$$p(A \text{ and } B) = p(A)p(B/A) = p(B)p(A/B). \tag{10.9}$$

The symbols $p(B/A)$ and $p(A/B)$ are referred to as *conditional probabilities.* The symbol $p(B/A)$ means, *the probability of B given that A has occurred.* The term *conditional probability* takes into account the possibility that the probability of B may depend on whether or not A has occurred, and conversely for $p(A/B)$.

Note: Either of the two events may be designated A or B since the symbols merely represent a convenient language for discussion, and there is no time order implied in the way they occur.

When events are independent. In the special case where the occurrence of A is in no way related to the occurrence of B and vice versa, the events are said to be *independent.* Independence is shown symbolically by $p(A/B) = p(A)$ and $p(B/A) = p(B)$. When events are *independent,* the multiplication rule simplifies to

$$p(A \text{ and } B) = p(A)p(B). \tag{10.10}$$

Let us take a look at a simple example. Imagine that we toss a pair of "honest" dice and that we want to determine the probability of obtaining 2 fours. Since the two events are independent (the outcome on one die does not affect the outcome on the second), we may employ the multiplication rule for independent events. Since the probability of obtaining a four is $\frac{1}{6}$, our answer becomes

$$p(A \text{ and } B) = (\tfrac{1}{6})(\tfrac{1}{6}) = \tfrac{1}{36}.$$

Suppose we flip a coin and toss a die, and we want to know the probability of obtaining a head on the coin and a five on the die. Since the probability of

obtaining a head is $\frac{1}{2}$, and that of obtaining a five is $\frac{1}{6}$, our answer is

$$p(A \text{ and } B) = (\tfrac{1}{2})(\tfrac{1}{6}) = \tfrac{1}{12}.$$

We should note, in passing, that mutually exclusive events are never independent, since the occurrence of one *denies* the possibility of the occurrence of a second.

Stated symbolically,

$$p(A/B) = p(B/A) = 0 \tag{10.11}$$

when events are *mutually exclusive.*

Thus, if we flip a single coin, the probability of obtaining both a head and a tail is zero.

When events are related (nonindependent). Let us look at an example to illustrate the application of the multiplication rule when events are not independent.

In a single throw of two dice, what is the probability that the sum will be an even number (event A) and that a 6 will appear on at least one of the dice

Table 10.1

All outcomes possible in a single throw of two dice (or two consecutive throws of a single die)

11	21	31	41	51	61
12	22	32	42	52	62
13	23	33	43	53	63
14	24	34	44	54	64
15	25	35	45	55	65
16	26	36	46	56	66

(event B)? All the possible outcomes are listed in Table 10.1. The first member of each pair designates the number appearing on one of the dice, and the second member of the pair designates the number appearing on the other die.

Since in the total of 36 outcomes, 18 sum to an even number (circled), it follows that $p(A) = \frac{18}{36}$. Similarly, since 11 of these 36 outcomes contain a six on at least one of the dice, $p(B) = \frac{11}{36}$. Of the 18 even sums, 5 sums have a six appearing on at least one of the dice. Thus, $p(B/A)$ is $\frac{5}{18}$, which is the prob-

ability that a six will appear on at least 1 of the dice, given the knowledge that the sum is an even number.

Similarly, given the knowledge that a six appears on at least one of the dice, the probability that the sum is even, $p(A/B)$, is $\frac{5}{11}$.

In summary, we have found that

$$p(A) = \tfrac{18}{36}, \qquad p(B) = \tfrac{11}{36},$$
$$p(B/A) = \tfrac{5}{18}, \qquad p(A/B) = \tfrac{5}{11}.$$

Employing formula (10.9), we obtain

$$p(A \text{ and } B) = p(A)p(B/A) = \tfrac{18}{36} \cdot \tfrac{5}{18} = 0.14,$$

or

$$p(A \text{ and } B) = p(B)p(A/B) = \tfrac{11}{36} \cdot \tfrac{5}{11} = 0.14.$$

10.4.5 Sampling with and without Replacement

If we were to select two cards from a well-shuffled deck of cards, what is the probability that both would be queens? There are two ways of selecting the cards: (1) We might select one card, replace it in the deck, reshuffle, and draw a second card. This procedure is known as *sampling with replacement.* (2) We might select two cards consecutively without replacing the first in the deck. This method is known as *sampling without replacement.*

Let A be the event of a queen on the first draw, and B the event of a queen on the second draw. Now, when we employ *sampling with replacement,* the probability of drawing a queen remains the same on both draws since we return the first card drawn. Thus, since $p(A/B) = p(A)$ and $p(B/A) = p(B)$, the two draws are independent. We may therefore use formula (11.10):

$$p(A \text{ and } B) = p(A)p(B) = \tfrac{1}{13} \cdot \tfrac{1}{13} = \tfrac{1}{169}.$$

When we employ *sampling without replacement,* the probability of obtaining a queen on the second draw is reduced whenever the first card selected is a queen. In other words, when $p(B/A) \neq p(B)$, or $p(A/B) \neq p(A)$, the events are *not* independent. The conditional probability of drawing a queen on the second trial is $\frac{3}{51}$. Thus, employing formula (11.9), we find that the probability of selecting two queens on consecutive draws from a deck of cards, without replacement, becomes

$$p(A \text{ and } B) = p(A)p(B/A) = (\tfrac{1}{13})(\tfrac{3}{51}) = \tfrac{1}{221}.$$

It should be noted that the difference between sampling with and without replacement is negligible when the population is large relative to the size of the

sample. Thus in public opinion polling, where the population may number in the millions, sampling without replacement may be employed without danger of doing violence to the basic assumptions of probability theory.

Finally, the above formulations may be generalized to include any number of events. For independent events, the formula is a simple generalization of the basic formula. Thus for three independent events, the joint probability is defined as follows:

$$p(A, B, \text{and } C) = p(A)p(B)p(C). \tag{10.12}$$

To illustrate: What is the probability of obtaining a queen, a head, and a 3 in one selection from a deck of cards, one flip of a coin, and one toss of a die?

$$p(A, B, \text{and } C) = (\tfrac{1}{13})(\tfrac{1}{2})(\tfrac{1}{6}) = \tfrac{1}{156}.$$

Thus far, we have developed a probability model for discrete variables. Needless to say, the treatment is not exhaustive. However, since much research data involve continuous scales of measurement, it is necessary to develop a probability model for continuous variables.

10.5 INTRODUCTION TO SET THEORY*

Imagine that we conduct a conceptual experiment involving the familiar children's game "guess what hand it is in." What are the various outcomes that are possible. The hidden object is in either the right hand or the left hand. These two outcomes delineate the entire *sample space* (all outcomes) of our conceptual experiment.

This experiment is referred to as a *simple experiment* since each outcome cannot be further broken down into component outcomes. Note that, on any given trial, only one of these outcomes may occur. Thus, in a simple experiment, the outcomes are mutually exclusive.

If we were to play the same game with an octopus, the sample space for a single trial would consist of eight possible outcomes. Again the experiment would be described as simple since each outcome in the sample space is irreducible.

Let us imagine that we extended our child's game to two trials, employing another person rather than an octopus as our playmate. The sample space would now be described by the following outcomes:

$$RR \qquad RL \qquad LR \qquad LL$$

Since each of these outcomes is reducible to the component outcomes resulting from each trial, the experiment is described as complex. Thus, the out-

* This section may be omitted without any loss in continuity.

come RR consists of hiding the object in the right hand on two successive trials. If we are not interested in the ordering of the outcomes (i.e., if we regard R on the first trial and L on the second as being the same as L on the first trial and R on the second), the two middle outcomes may be combined into one and we may speak of four different ways of obtaining the three outcomes in our conceptual experiment.

Let us look at one additional example. Imagine that we have four different score values: 1, 2, 3, and 4. Drawing samples of $n = 3$, without replacement and without regard to ordering, would yield the following sample space:

1, 2, 3
1, 2, 4
1, 3, 4
2, 3, 4

If we sampled without replacement but wished to include all different orderings of outcomes, we would obtain the sample space presented in Table 10.2.

Table 10.2 Sample space obtained by drawing samples $n = 3$, without replacement, from a population of four members

1, 2, 3	2, 1, 3	3, 1, 2	4, 1, 2
1, 3, 2	2, 3, 1	3, 2, 1	4, 2, 1
1, 2, 4	2, 1, 4	3, 1, 4	4, 1, 3
1, 4, 2	2, 4, 1	3, 4, 1	4, 3, 1
1, 3, 4	2, 3, 4	3, 2, 4	4, 2, 3
1, 4, 3	2, 4, 3	3, 4, 2	4, 3, 2

A *set* or an *event* may be defined as a collection of outcomes. If we were interested in the event: all samples which yield the numerical value 1 in the second position, we would find that six different outcomes comprise the set. It is important to note that an event may also include the null case (referred to as the *null set* or $\emptyset$) which occurs when the collection of outcomes comprising the event does not appear in the sample space. Thus, the events $\overline{X} < 2$ or $\overline{X} > 3$ do not appear in the sample space. Similarly, the event may also include all outcomes, e.g., $2 \leq \overline{X} \leq 3$.

10.5.1 Set Operations

Let us define set or event A as any combination of three numbers in the sample space shown in Table 10.2 which sums to six. We see that there are six members which make up this set, viz. (1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), and (3, 2, 1).

Let us define a second set, B, as any combination of three numbers in Table 10.2 which has a 3 in the second position. This set consists of the following groupings: (1, 3, 2), (1, 3, 4), (2, 3, 1), (2, 3, 4), (4, 3, 1), and (4, 3, 2).

Finally, let us define a third event or set, C, as all groupings of three numbers which contain a 4 in the third position. This set contains the following members (1, 2, 4), (1, 3, 4), (2, 1, 4), (2, 3, 4), (3, 1, 4), and (3, 2, 4).

10.5.2 Combining of Sets

In a procedure analogous to addition of numbers, two or more sets or events may be combined by an operation referred to as *union* of sets. The union of two sets consists of combining those members which are in *either* set as well as those that are in *both*. The union of sets is shown symbolically by the "cup," e.g., $A \cup B$, $A \cup C$, $B \cup C$, $A \cup B \cup C$.

The union of sets may be shown visually by the use of Venn diagrams (Fig. 10.1)

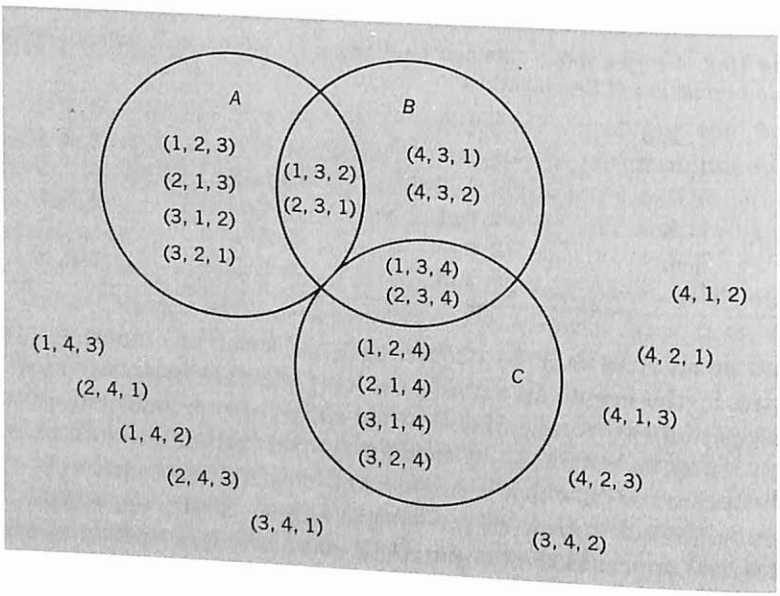


Fig. 10.1 Venn diagram showing the sample space obtained by drawing samples $n = 3$, without replacement, from a population of four numbers. The sets A, B, and C are defined as follows:

A = any combination of three numbers in the sample space which sums to six.
B = any combination of three numbers in the sample space which has a 3 in the second position.
C = any combination of three numbers in the sample space which has a 4 in the third position.

The union of $A \cup B$, $A \cup C$, $B \cup C$ are shown as follows:

$$\begin{aligned}
A \cup B = {} & (1, 2, 3),\ (2, 1, 3),\ (3, 1, 2),\ (3, 2, 1),\ (1, 3, 2),\ (2, 3, 1),\ (4, 3, 1),\\
& (4, 3, 2),\ (1, 3, 4),\ (2, 3, 4)\\
A \cup C = {} & (1, 2, 3),\ (2, 1, 3),\ (3, 1, 2),\ (3, 2, 1),\ (1, 3, 2),\ (2, 3, 1),\ (1, 3, 4),\\
& (2, 3, 4),\ (1, 2, 4),\ (2, 1, 4),\ (3, 1, 4),\ (3, 2, 4)\\
B \cup C = {} & (1, 3, 2),\ (2, 3, 1),\ (4, 3, 1),\ (4, 3, 2),\ (1, 3, 4),\ (2, 3, 4),\ (1, 2, 4),\\
& (2, 1, 4),\ (3, 1, 4),\ (3, 2, 4)
\end{aligned}$$

Note that each of the sets $A \cup B$ and $B \cup C$ contains ten elements although A, B, and C each contains six elements. This is due to the fact that $A \cup B$ and $B \cup C$ each contains common elements which are counted only once. On the other hand, sets A and C do not contain common elements. Therefore $A \cup C$ contains a total of twelve elements.

The *intersection* of two sets denotes the elements of these sets which are shared in common. Thus, the intersection of A and B, symbolized by $\cap$, represents all possible events which are common to sets A and B.

Thus,

$$A \cap B = (1, 3, 2),\ (2, 3, 1).$$

Similarly,

$$B \cap C = (1, 3, 4),\ (2, 3, 4).$$

The intersection of A and C, which have no elements in common, produces the *null set*, $\emptyset$. Thus, $A \cap C = \emptyset$.

When two or more sets do not overlap, e.g., when $A \cap C = \emptyset$, they are said to be *mutually exclusive*.

We may now state the probability rules appearing in Section 10.4 in terms of set theory.

Addition rule 1:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \tag{10.13}$$

There are a total of twenty-four elements comprising the sample space shown in Table 10.2. There are six elements in each of the sets A and B.

Thus, $p(A) = \frac{6}{24}$ and $p(B) = \frac{6}{24}$.

The intersection of A and B $(A \cap B)$ contains two elements. Thus, $p(A \cap B) = \frac{2}{24}$.

Therefore,

$$p(A \cup B) = \tfrac{6}{24} + \tfrac{6}{24} - \tfrac{2}{24} = \tfrac{10}{24} = \tfrac{5}{12}$$

Addition rule 2: If the intersection of two sets produces a null set, i.e., the events are *mutually exclusive*, the addition rule reduces to:

$$p(A \cup C) = p(A) + p(C), \tag{10.14}$$

where $p(A \cap C) = 0$.

Thus, in the sample space shown in Table 10.2, the probability of obtaining the event A or C is given by:

$$\begin{aligned} p(A \cup C) &= p(A) + p(C) \\ &= \tfrac{6}{24} + \tfrac{6}{24} = \tfrac{12}{24} \\ &= \tfrac{1}{2}. \end{aligned}$$

Addition rule 3: The addition rule for determining $p(A \cup B \cup C)$ may be generalized as follows:

$$p(A \cup B \cup C) = p(A) + p(B) + p(C) - [p(A \cap B) + p(A \cap C) + p(B \cap C)] \tag{10.15}$$

In the present example, $p(A) = \frac{6}{24}$, $p(B) = \frac{6}{24}$, $p(C) = \frac{6}{24}$, $p(A \cap B) = \frac{2}{24}$, $p(A \cap C) = 0$, and $p(B \cap C) = \frac{2}{24}$.

Thus, $\begin{aligned}[t] p(A \cup B \cup C) &= \tfrac{6}{24} + \tfrac{6}{24} + \tfrac{6}{24} - (\tfrac{2}{24} + 0 + \tfrac{2}{24}) \\ &= \tfrac{18}{24} - \tfrac{4}{24} = \tfrac{14}{24} \\ &= \tfrac{7}{12}. \end{aligned}$

Multiplication rule 1: If two events are *mutually exclusive* (i.e., they do not overlap), the probability of their joint occurrence is

$$p(A \cap C) = 0.$$

Thus, for the sample space given in Table 10.2, the probability of the joint occurrence of the events A and C is zero.

Multiplication rule 2: If two events are not mutually exclusive (i.e., they overlap), the probability of their joint occurrence is given by the probability of the intersection of the two sets.

Thus, $p(A \cap B)$, in the present problem, is $\frac{2}{24}$ or $\frac{1}{12}$. Similarly, $p(B \cap C) = \frac{1}{12}$.

10.6 PROBABILITY AND CONTINUOUS VARIABLES

Up to this point we have considered probability in terms of the expected relative frequency of an event. In fact, as you will recall, probability was defined in terms of frequency and expressed as the following proportion [formula (10.1)]:

$$p(A) = \frac{\text{no. outcomes favoring event } A}{\text{total no. outcomes}}.$$

However, this definition presents a problem when we are dealing with continuous variables. As we pointed out in Section 3.7.1, it is generally advisable to represent frequency in terms of areas under a curve when we are dealing with continuous variables. Thus, for continuous variables, we may express probability as the following proportion:

$$p = \frac{\text{area under portions of a curve}}{\text{total area under the curve}}. \tag{10.16}$$

Since the total area in a probability distribution is equal to 1.00, we define p as the proportion of total area under portions of a curve.

Chapters 11 through 15 employ the standard normal curve as the probability model. Let us examine the probability-area relationship in terms of this model.

10.7 PROBABILITY AND THE NORMAL CURVE MODEL

In Section 7.3 we stated that the standard normal distribution has a μ of 0, a σ of 1, and a total area that is equal to 1.00. We saw that when scores on a normally distributed variable are transformed into z-scores, we are, in effect, expressing these scores in units of the standard normal curve. This permits us to express the difference between any two scores as proportions of total area under the curve. Thus we may establish probability values in terms of these proportions as in formula (10.16).

Let us look at several examples which illustrate the application of probability concepts to the normal curve model.

10.7.1 Illustrative Problems*

For all problems, assume $\mu = 100$ and $\sigma = 16$.

Problem 1

What is the probability of selecting at *random*, from the general population, a person with an I.Q. score of at least 132? The answer to this question is given by the proportion of area under the curve above a score of 132 (Fig. 10.2).

First, we must find the z-score corresponding to $X = 132$.

$$z = \frac{132 - 100}{16} = 2.00.$$

In Column C (Table A), we find that 0.0228 of the area lies at or beyond a z of 2.00. Therefore, the probability of selecting, at random, a score of at least 132 is 0.0228.

* See Section 7.4.

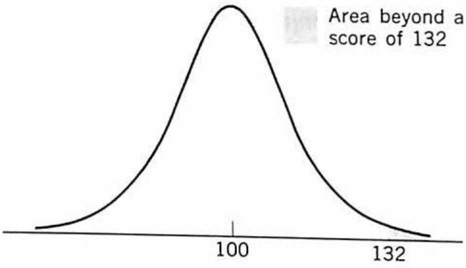


Fig. 10.2 Proportion of area above a score of 132 in a normal distribution with $\mu = 100$ and $\sigma = 16$.

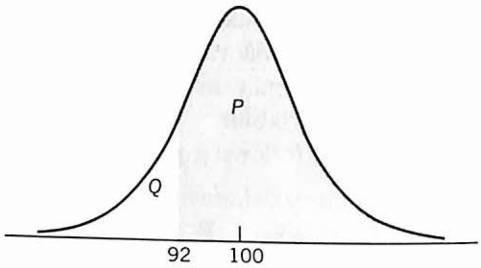


Fig. 10.3 Proportion of area above (P) and below (Q) a score of 92 in a normal distribution with $\mu = 100$ and $\sigma = 16$.

Problem 2

What is the probability of selecting, at random, an individual with an I.Q. score of at least 92?

We are dealing with two mutually exclusive and exhaustive areas under the curve. The area under the curve above a score of 92 is P; the area below a score of 92 is Q. In solving our problem, we may therefore employ formula (10.7):

$$P = 1.00 - Q.$$

By expressing a score of 92 in terms of its corresponding z, we may obtain the proportion of area below $X = 92$ (that is Q) directly from Column C (Table A).

The z-score corresponding to $X = 92$ is

$$z = \frac{92 - 100}{16} = -0.50.$$

The proportion of area below a z of -0.50 is 0.3085. Therefore, the probability of selecting, at random, a score of at least 92 becomes

$$P = 1.00 - 0.3085 = 0.6915.$$

Figure 10.3 illustrates this relationship.

Problem 3

Let us look at an example involving the multiplication law. Given that sampling with replacement is employed, what is the probability of drawing, at random, three individuals with I.Q.'s equaling or exceeding 124? For this problem, we will again assume that $\mu = 100$ and $\sigma = 16$.

The z-score corresponding to $X = 124$ is

$$z = \frac{124 - 100}{16} = 1.5,$$

In Column C (Table A), we find that 0.1336 of the area lies at or beyond $X = 124$. Therefore,

$$p(A, B, C) = (0.1336)^3 = 0.0024.$$

10.8 ONE- AND TWO-TAILED p-VALUES

In Problem 1 we posed the question: What is the probability of selecting a person with a score as high as 132? We answered the question by examining only one tail of the distribution, namely, scores as high as or higher than 132. For this reason, we refer to the probability value that we obtained as being a *one-tailed p-value.*

In statistics and research, the following question is more commonly asked: "What is the probability of obtaining a score (or statistic) this *deviant* from the mean? . . . or a score (or statistic) this *rare* . . . or a result this *unusual?*" Clearly, when the frequency distribution of scores is symmetrical, a score of 68 or lower is every bit as deviant from a mean of 100 as a score of 132. That is, both are two standard deviation units away from the mean. When we express the probability value, taking into account both tails of the distribution, we refer to the p-value as being *two-tailed.* In symmetrical distributions, two-tailed p-values may be obtained merely by doubling the one-tailed probability value. Thus in the preceding problem, the probability of selecting a person with a score as *rare* or *unusual* as 132 is $2 \times 0.0228 = 0.0456$.

The distinction between one- and two-tailed probability values takes on added significance as we progress into inferential statistics.

CHAPTER SUMMARY

In this chapter we discussed:

1. The importance of the concept of randomness in inferential statistics. The basic characteristic of random events is known as independence. Although the individual events are unpredictable, collections of random events take on characteristic and predictable forms. The binomial distribution and the normal curve were cited in this regard.

2. The theory of probability which is concerned with the outcomes of experiments. We can distinguish between probabilities established by assuming *idealized* relative frequencies, and those established empirically by determining

relative frequencies. Probability was defined as

$$p = \frac{\text{no. of outcomes favoring event}}{\text{total no. of outcomes}}.$$

3. The formal properties of probability:
   a) *Probabilities vary between* 0 *and* 1.00
   b) *The addition rule*
   If A and B are two events, the probability of obtaining either of them is equal to the probability of A plus the probability of B minus the probability of their joint occurrence. Thus

   $$p(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B).$$

   If events A and B are *mutually exclusive*, the addition rule becomes

   $$p(A \text{ or } B) = p(A) + p(B).$$

   If the two events are *mutually exclusive and exhaustive*, we obtain

   $$p(A) + p(B) = 1.00.$$

   Allowing P to represent the probability of occurrence and Q to represent the probability of nonoccurrence, we find that three additional useful formulations for mutually exclusive and exhaustive events are

   $$P + Q = 1.00, \qquad P = 1.00 - Q, \qquad Q = 1.00 - P.$$

   c) *The multiplication rule*
   Given two events, A and B, the probability of obtaining both A and B jointly is the product of the probability of obtaining one of these events times the conditional probability of obtaining one event, given that the other event has occurred. Thus

   $$p(A \text{ and } B) = p(A)p(B/A) = p(B)p(A/B).$$

   When the events are *independent*, this formulation becomes

   $$p(A \text{ and } B) = p(A)p(B).$$

4. An optional section on set theory and its relationship to probability theory was presented.

5. In the application of probability theory to continuously distributed variables, probability is expressed in terms of the proportion of area under a curve.

Hence
$$p = \frac{\text{area under portions of a curve}}{\text{total area under a curve}}.$$

We saw how we may employ z-scores and the standard normal curve to establish various probabilities for normally distributed variables.

6. Finally, we distinguished between one- and two-tailed probability values.

Terms to Remember:

Random
Independence
Bias
Binomial or Bernoulli model
Sampling distribution
Probability
Mutually exclusive
Exhaustive
Joint occurrence
Conditional probability
Sampling with replacement
Sampling without replacement
Addition rule
Multiplication rule
Sample space
Set
Event
Null set
Union
Venn diagram
Intersection
One-tailed p-values
Two-tailed p-values
Deviant, rare, unusual

EXERCISES

1. List all the possible outcomes of a coin that is tossed three times. Calculate the probability of
   a) 3 heads,
   b) 3 tails,
   c) 2 heads and 1 tail,
   d) at least 2 heads.
2. A card is drawn at random from a deck of 52 playing cards. What is the probability that
   a) it will be the ace of spaces?
   b) it will be an ace?
   c) it will be an ace or a face card?
   d) it will be a spade or a face card?
3. Express the probabilities, in Problems 1 and 2, in terms of *odds against*.
4. In a single throw of two dice, what is the probability that
   a) a 7 will appear?
   b) a doublet (two of the same number) will appear?
   c) a doublet or an 8 will appear?
   d) an even number will appear?
5. On a slot machine (commonly referred to as a "one-armed bandit"), there are three reels with five different fruits plus a star on each reel. After inserting a coin and pulling the handle, the player sees that the three reels revolve independently

several times before stopping. What is the probability that
a) three lemons will appear?
b) any three of a kind will appear?
c) two lemons and a star will appear?
d) two lemons and any other fruit will appear?
e) no star will appear?

6. Three cards are drawn at random (without replacement) from a deck of 52 cards. What is the probability that
   a) all three will be hearts?
   b) none of the three cards will be hearts?
   c) all three will be face cards?
7. Calculate the probabilities in Problem 6 if each card is replaced after it is drawn.
8. A well-known test of intelligence has a mean of 100 and a standard deviation of 16.
   a) What is the probability that someone picked at random will have an I.Q. of 122 or higher?
   b) There are I.Q.'s so *high* that the probability is 0.05 that such I.Q.'s would occur in a random sample of people. Those I.Q.'s are beyond what value?
   c) There are I.Q.'s so *extreme* that the probability is 0.05 that such I.Q.'s would occur in a random sample of people. Those I.Q.'s are beyond what values?
   d) The next time you shop you will undoubtedly see someone who is a complete stranger to you. What is the probability that his I.Q. will be between 90 and 110?
   e) What is the probability of selecting two people, at random,
      i) with I.Q.'s of 122 or higher?
      ii) with I.Q.'s between 90 and 110?
      iii) one with an I.Q. of 122 or higher, the other with an I.Q. between 90–110?
   f) What is the probability that on leaving your class, the first student you meet will have an I.Q. below 120? Can you answer this question on the basis of the information provided above? If not, why not?
9. Which of the following selection techniques will result in random samples? Explain your answers.
   a) Population: Viewers of a given television program. Sampling technique: On a given night, interviewing every fifth person in the studio audience.
   b) Population: A home-made pie. Sampling technique: A wedge selected from any portion of the pie.
   c) Population: All the children in a suburban high school. Sampling technique: Selecting one child sent to you by each homeroom teacher.
10. In a study involving a test of visual acuity, four different hues varying slightly in brightness are presented to the subject. What is the probability that he will arrange them in order, from greatest brightness to least, by chance.
11. The proportion of people with type A blood in a particular city is 0.20. What is the probability that:
    a) A given individual, selected at random, will have type A blood?
    b) Two out of two individuals will have type A blood?

c) A given individual will *not* have type A blood?
d) Two out of two individuals will *not* have type A blood?

12. In a manufacturing process, the proportion of items that are defective is 0.10. What is the probability that:
    a) In a sample of four items, none will be defective?
    b) In a sample of four items, all will be defective?
    c) One or more but less than four will be defective?
13. In the manufacture of machine screws for the space industry, millions of screws measuring 0.010 cm are produced daily. The standard deviation is 0.001. A screw is considered defective if it deviates from 0.010 by as much as 0.002. What is the probability that:
    a) One screw, selected at random, will be defective?
    b) Two out of two screws will be defective?
    c) One screw, selected at random, will *not* be defective?
    d) Two out of two screws will *not* be defective?
    e) One screw, selected at random, will be too large?
    f) Two out of two screws will be too small?
14. A bag contains 6 blue marbles, 4 red marbles, and 2 green marbles. If you select a single marble at random from the bag, what is the probability that it will be:
    a) red? b) blue? c) green? d) white?
15. Selecting *without* replacement from the bag described in Problem 14, what is the probability that:
    a) Three out of three will be blue?
    b) Two out of two will be green?
    c) None out of four will be red?
16. Selecting *with* replacement from the bag described in Problem 14, what is the probability that:
    a) Three out of three will be blue?
    b) Two out of two will be green?
    c) None out of four will be red?
17. Forty percent of the students at a given college major in business administration. Seventy percent of these are male and thirty percent female. Sixty percent of the students in the school are male. What is the probability that:
    a) One student, selected at random, will be a BA major.
    b) One female, selected at random, will be a BA major.
    c) Two students, selected at random, will both be BA majors.
    d) One male and one female, selected at random, will both be BA majors.
18. What is the probability that a score chosen at random from a normally distributed population with a mean of 66 and a standard deviation of 8 will be:
    a) greater than 70?
    b) less than 60?
    c) between 60 and 70?
    d) in the 70's?
    e) either less than 55 or greater than 72?

f) either less than 52 or between 78 and 84?
g) either between 56 and 64 or between 80 and 86?

19. What is meant by random sampling?

20. Construct the sample space resulting from a conceptual experiment involving the tossing of three coins simultaneously (or one coin successively).
 a) Prepare a Venn diagram showing the events: Event A, at least two heads; Event B, two or one heads. Find: $p(A \cup B)$; $p(A \cap B)$.
 b) Prepare a Venn diagram showing the events: Event A, three heads; Event B, two or fewer heads. Find: $p(A \cup B)$; $p(A \cap B)$.

21. Construct the sample space resulting from a conceptual experiment involving the tossing of two coins and one die simultaneously (or two coins and one die successively).
 a) Prepare a Venn diagram showing the events: Event A, two heads and a 6; Event B; one head and a 5. Find: $p(A \cup B)$; $p(A \cap B)$.
 b) Prepare a Venn diagram showing the events: Event A, at least one head and a 5; Event B, two heads and a 5. Find: $p(A \cup B)$; $p(A \cap B)$.

Introduction to Statistical Inference* 11

11.1 WHY SAMPLE?

You are the leader of a religious denomination, and for the purpose of planning recruitment you want to know what proportion of the adults in the United States claim church membership. How would you go about getting this information?

You are a rat psychologist and you are interested in the relationship between strength of drive and learning. Specifically, what are the effects of duration of hunger drive on the number of trials required for a rat to learn a T-maze?

You are a sociologist and you want to study the differences in child rearing practices among parents of delinquent versus nondelinquent children.

You are a market researcher and you want to know what proportion of individuals prefer different car colors and their various combinations.

You are a park attendant and you want to determine whether the ice is sufficiently thick to permit safe skating.

You are a gambler and you want to determine whether a set of dice is "dishonest."

What do each of these problems have in common? You are asking questions about the parameter of a population to which you want to generalize your answers, but you have no hope of ever studying the *entire* population. Earlier (Section 1.2) we defined a population as a *complete* set of individuals, objects, or measurements having some common observable characteristic. It is frequently impossible to study *all* the members of a given population either because the population as defined has an infinite number of members, or because

* It is recommended that you reread Section 1.2 for purposes of reviewing several definitions of terms that will appear in this chapter.

the population is so large that it defies exhaustive study. Consequently, when we refer to *the population* we are often dealing with a hypothetical entity.†

For example, the park attendant could not possibly determine the thickness of the ice at all points. Therefore, he must rely on a relatively small sample of measurements regarding the thickness of the ice. The religious leader, as well, could not reasonably hope to obtain replies from every adult in the United States.

Since populations can rarely be studied exhaustively, we must depend on samples as a basis for arriving at a hypothesis concerning various characteristics, or parameters, of the population. Note that our interest is not in descriptive statistics, *per se*, but in making inferences from data. Thus, if we ask 100 people how they intend to vote in a forthcoming election, our primary interest is not how these 100 people will vote, but in estimating how the entire voting population will cast their ballots.

Almost all research involves the observation and the measurement of a limited number of individuals or events. These measurements are presumed to tell us something about the population. In order to understand how we are able to make inferences about a population from a sample, it is necessary to introduce the concept of sampling distributions.

11.2 THE CONCEPT OF SAMPLING DISTRIBUTIONS

In actual practice, inferences about the parameters of a population are made from statistics that are calculated from a sample of N observations drawn at random from this population. If we continued to draw samples of size N from this population, we should not be surprised if we found some differences among the values of the sample statistics obtained. Indeed, it is this observation that has led to the concept of sampling distributions.

> *A sampling distribution is a theoretical probability distribution of the possible values of some sample statistic which would occur if we were to draw all possible samples of a fixed size from a given population.*

Imagine that we have a population of scores. We decide to draw samples of three at a time and calculate the mean and the standard deviation of each sample. If we were to draw all the possible samples in which $N = 3$, and plot the resulting means and the standard deviations, we would have two sampling distributions: one for the means and one for the standard deviations.

Imagine that we have a coin and toss it an unlimited series of trials, with ten tosses per trial. After each trial, we record the proportion of heads. If we were

† In the typical experimental situation, the actual population, or universe, does not exist. What we attempt to do is to find out something about the characteristics of that population *if it did* exist. Thus our sample groups provide us with information about the characteristics of a population if it did, in fact, exist.

to obtain a large number of samples in this way and were to construct a distribution showing the frequency with which each proportion was obtained, we would approximate the sampling distribution of a two-category variable for $N = 10$.

Why is the concept of a sampling distribution so important? The answer is simple. Whenever we estimate a population parameter from a sample, we shall ask questions such as: "How good an estimate do I have? Can I conclude that the population parameter is identical with the sample statistic? Or is there likely to be some error? If so, how much?" To answer each of these questions, we will compare our sample results with the "expected" results. The expected results are, in turn, given by the appropriate sampling distribution. But, what does the sampling distribution of a particular statistic look like? How can we ever know the form of the distribution, and thus, what the expected results are? Since the inferences we will be making imply knowledge of the *form* of the sampling distribution, it is necessary to set up certain idealized *models*. The normal curve and the binomial distribution are two models whose mathematical properties are known. Consequently, these two distributions are frequently employed as models to describe particular sampling distributions. Thus, for example, if we know that the sampling distribution of a particular statistic takes the form of a normal distribution, we may use the known properties of the normal distribution to make inferences and predictions about the statistic.

The following sections should serve to clarify these important points.

11.3 BINOMIAL DISTRIBUTION

Let us say that you have a favorite coin which you use constantly in every day life as a basis of "either-or" decision making. For example, you may ask, "Should I study tonight for the statistics quiz, or should I relax at one of the local movie houses? Heads, I study, tails, I don't." Over a period of time, you have sensed that the decision has more often gone "against you" than "for you" (in other words, you have to study more often than relax!). You begin to question the accuracy and the adequacy of the coin. Does the coin come up heads more often than tails? How might you find out?

One thing is clear. The true proportion of heads and tails characteristic of this coin can never be known. You could start tossing the coin this very minute and continue for a million years (granting a long life and a remarkably durable coin) and you would not exhaust the population of possible outcomes. In this instance, the true proportion of heads and tails is unknowable because the universe, or population, is unlimited.

The fact that the *true* value is unknowable does not prevent us from trying to estimate what it is. We have already pointed out that since populations can rarely be studied exhaustively, we must depend on samples to estimate the parameters.

Returning to our problem with the coin, we clearly see that in order to determine whether or not the coin is biased, we will have to obtain a sample of the "behavior" of that coin and arrive at some generalization concerning its possible bias.

Let us define an *unbiased* coin as one in which the probability of heads is equal to the probability of tails. We may employ the symbol P to represent the probability of the occurrence of a head, and Q, the probability of the nonoccurrence of a head (i.e., the occurrence of a tail). Since we are dealing with two mutually exclusive and exhaustive outcomes, if $P = Q = \frac{1}{2}$, the coin is *unbiased*. Conversely, if $P \neq Q \neq \frac{1}{2}$, the coin is *biased*.

How do we determine whether a particular coin is biased or unbiased? Suppose we conduct the following experiment. We toss the coin 10 times and obtain 9 heads and 1 tail. This may be viewed as a sample of the "behavior" of this coin. On the basis of this result, are we now justified in concluding that the coin is biased? Or, is it reasonable to expect as many as 9 heads from a coin that is unbiased? Before we answer these questions, it is necessary to look at the sampling distribution of *all* possible outcomes. Let us see how we might construct this sampling distribution, employing a hypothetical coin.

11.3.1 Construction of Binomial Sampling Distributions by Enumeration

First, we must assume that this coin is unbiased, and that there are only two possible outcomes resulting from each toss of the coin: heads or tails. It will not stand on its side; it will not become lost; it cannot turn up both heads and tails at the same time.

If we toss this coin twice ($N = 2$), there are four possible ways the coin may fall: HH, HT, TH, and TT. The two middle ways may be thought of as the same outcome in that each represents one head and one tail. Thus tossing an *unbiased* coin twice results in the following theoretical frequency distribution:

Outcome	No. of ways for specified outcome to occur
$2H$	1
$1H, 1T$	2
$2T$	1
	4

Note that in N tosses of a coin, there are $N + 1$ different possible outcomes and 2^N different ways to obtain these $N + 1$ outcomes. Thus, when $N = 2$, there are four ways of obtaining the three different outcomes.

We may calculate the probability associated with each outcome by dividing the number of ways each outcome may occur by 2^N. For example, using for-

Table 11.1

All possible outcomes obtained by tossing an unbiased coin five times ($N = 5$)

Number of heads					
$0H$	$1H$	$2H$	$3H$	$4H$	$5H$
$T\,T\,T\,T\,T$	$H\,T\,T\,T\,T$	$H\,H\,T\,T\,T$	$H\,H\,H\,T\,T$	$H\,H\,H\,H\,T$	$H\,H\,H\,H\,H$
	$T\,H\,T\,T\,T$	$H\,T\,H\,T\,T$	$H\,H\,T\,H\,T$	$H\,H\,H\,T\,H$	
	$T\,T\,H\,T\,T$	$H\,T\,T\,H\,T$	$H\,T\,H\,H\,T$	$H\,H\,T\,H\,H$	
	$T\,T\,T\,H\,T$	$H\,T\,T\,T\,H$	$T\,H\,H\,H\,T$	$H\,T\,H\,H\,H$	
	$T\,T\,T\,T\,H$	$T\,H\,H\,T\,T$	$H\,H\,T\,T\,H$	$T\,H\,H\,H\,H$	
		$T\,H\,T\,H\,T$	$H\,T\,H\,T\,H$		
		$T\,H\,T\,T\,H$	$T\,H\,H\,T\,H$		
		$T\,T\,H\,H\,T$	$H\,T\,T\,H\,H$		
		$T\,T\,H\,T\,H$	$T\,H\,T\,H\,H$		
		$T\,T\,T\,H\,H$	$T\,T\,H\,H\,H$		

mula (10.1), the probability of obtaining one head and one tail in 2 tosses of an unbiased coin is

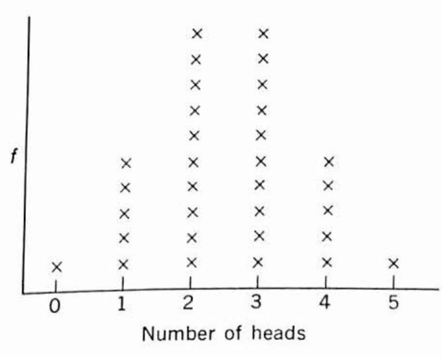
$$p(1H,1T) = \tfrac{2}{4} = 0.50.$$

We have seen how we can enumerate all the possible outcomes of tossing a hypothetical coin when $N = 2$, and construct corresponding frequency and probability distributions. This probability distribution represents the sampling distribution of outcomes when $N = 2$. Similarly, we may enumerate all the possible outcomes for any number of tosses of our hypothetical coin ($N = 3$, $N = 4$, etc.) and then construct the corresponding frequency and probability distributions.

Let us illustrate the construction of a probability distribution when $N = 5$. First, all possible outcomes are enumerated, as in Table 11.1. When $N = 5$, there are 32 ways (2^5) of obtaining the six different outcomes ($5 + 1$).

Fig. 11.1

Theoretical frequency distribution of various numbers of heads obtained by tossing an unbiased coin five times ($N = 5$).



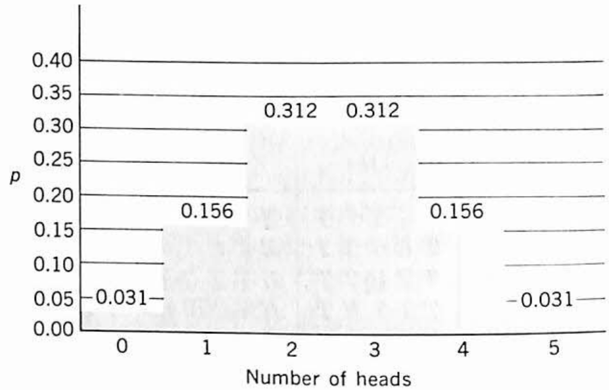


Fig. 11.2
Histogram of theoretical probability distribution of various numbers of heads obtained by tossing an unbiased coin five times ($N = 5$).

By placing the six different outcomes along the baseline and representing their frequency of occurrence along the ordinate, we have constructed the theoretical frequency distribution for the various outcomes when $N = 5$.

It is now possible to calculate the probability associated with each outcome. For example, using formula (10.1), the probability of obtaining 4 heads and 1 tail in 5 tosses of an unbiased coin is

$$p(4H, 1T) = \tfrac{5}{32} = 0.156.$$

A histogram of the complete probability distribution when $N = 5$ is shown in Fig. 11.2.

11.3.2 Construction of Binomial Sampling Distributions Employing the Binomial Expansion

Thus far, in order to calculate the probability associated with each outcome, we have had to enumerate all the possible ways in which various outcomes occur (see Table 11.1). As N increases, the process of enumeration becomes exceedingly laborious since the number of ways for the outcomes to occur (2^N) doubles with each additional toss of the coin.*

An alternative method of obtaining the sampling distribution of probabilities for a population consisting of two mutually exclusive and exhaustive categories ($P + Q = 1.00$) is given by the binomial expansion.

* In addition, when $P \neq Q \neq \frac{1}{2}$, it is not possible to obtain the probabilities associated with the various outcomes by simple enumeration. It is necessary to employ the binomial expansion to obtain these probabilities. Applications of the binomial expansion when $P \neq Q \neq \frac{1}{2}$ are presented in Chapter 17.

Table 11.2
Illustration of the binomial expansion when $P = Q = \frac{1}{2}$ and $N = 5$

	All heads	4H	3H	2H	1H	No heads
$(P+Q)^5 =$	P^5	$+ \frac{N}{1} P^4Q$	$+ \frac{N(N-1)}{(1)(2)} P^3Q^2$	$+ \frac{N(N-1)(N-2)}{(1)(2)(3)} P^2Q^3$	$+ \frac{N(N-1)(N-2)(N-3)}{(1)(2)(3)(4)} PQ^4$	$+ Q^5$
$=$	$(\frac{1}{2})^5$	$+ 5(\frac{1}{2})^4(\frac{1}{2})$	$+ 10(\frac{1}{2})^3(\frac{1}{2})^2$	$+ 10(\frac{1}{2})^2(\frac{1}{2})^3$	$+ 5(\frac{1}{2})(\frac{1}{2})^4$	$+ (\frac{1}{2})^5$
$=$	$\frac{1}{32}$	$+ \frac{5}{32}$	$+ \frac{10}{32}$	$+ \frac{10}{32}$	$+ \frac{5}{32}$	$+ \frac{1}{32}$
$=$	0.031	+ 0.156	+ 0.312	+ 0.312	+ 0.156	+ 0.031

Formula (11.1) presents the binomial expansion in its general form:

$$(P + Q)^N = P^N + \frac{N}{1} P^{N-1}Q + \frac{N(N-1)}{(1)(2)} P^{N-2}Q^2 + \frac{N(N-1)(N-2)}{(1)(2)(3)} P^{N-3}Q^3 + \cdots + Q^N. \qquad (11.1)$$

There are $N + 1$ terms to the right of the above equation, each representing a different possible outcome. The first term on the right-hand side of the equation (P^N) provides the probability of all events in the P-category; the second term provides the probability of all events in the P-category, except one; and, finally, the last term (Q^N) is the probability of all events in the Q-category.

To illustrate the binomial expansion, let us return to our preceding example involving 5 tosses of an unbiased coin. Since our hypothetical coin is unbiased, $P = Q = \frac{1}{2}$. When $N = 5$, there are $N + 1$ or six possible outcomes. These are shown in Table 11.2.

The numerators of the fractions in Table 11.2 (1, 5, 10, 10, 5, 1) are commonly referred to as the *coefficients* of the binomial expansion. These coefficients correspond exactly with the expected frequency of occurrence of each outcome, as previously shown in Fig. 11.1. In addition, the probabilities calculated from the binomial expansion agree exactly with those which we previously calculated (see Fig. 11.2).

11.4 TESTING STATISTICAL HYPOTHESES: LEVEL OF SIGNIFICANCE

At this point, you may wonder what happened to the experiment we were about to perform to determine whether our "decision-making" coin was biased. Having learned to calculate probability values, let us now address ourselves to the experiment. We are going to toss the coin a given number of times and determine whether or not the outcome is within certain expected limits. For ex-

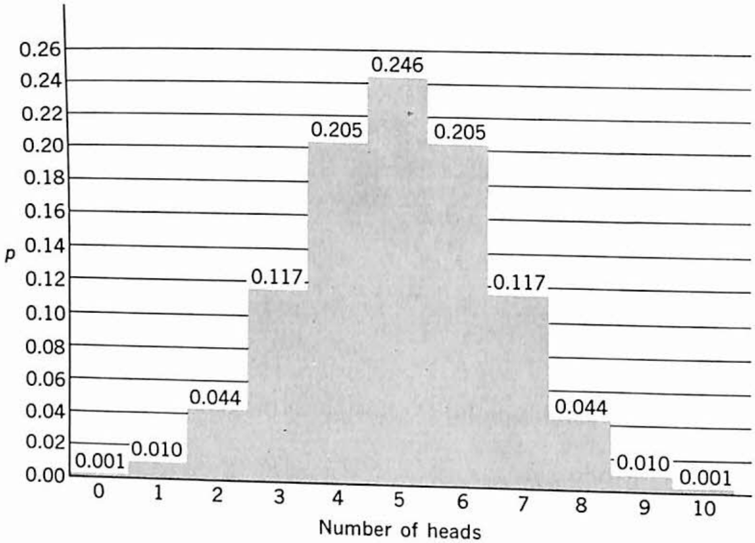


Fig. 11.3 Histogram of theoretical probability distribution of various numbers of heads obtained by tossing an unbiased coin ten times ($N = 10$).

ample, if we toss our coin 10 times and obtain 5 heads and 5 tails, would we begin to suspect our coin of being biased? Of course not, since this outcome is exactly a 50-50 split, and is in agreement with the hypothesis that the coin is not biased. What if we obtained 6 heads and 4 tails? Again, this is not an unusual outcome. In fact, if we expand the binomial, we can answer the question, "Given a theoretically perfect coin, how often would we expect an outcome at least this much different from a 50–50 split?" Reference to Fig. 11.3, which represents the theoretical probability distribution of various numbers of heads when $N = 10$, reveals that departures from a 50-50 split are quite common. Indeed, whenever we obtain either 6 or more heads, or 4 or fewer heads, we are departing from a 50-50 split. Such departures will occur fully 75.4% of the time when we toss a perfect coin in a series of trials with 10 tosses per trial.

What if we obtained 9 heads and 1 tail? Clearly, we begin to suspect the honesty of the coin. Why? At what point do we change from attitudes accepting the honesty of the coin to attitudes rejecting its honesty? This question takes us to the crux of the problem of inferential statistics. We have seen that the more unusual or rare the event, the more prone we are to look for nonchance explanations of the event. When we obtained 6 heads in 10 tosses of our coin, we felt no necessity to find an explanation for its departure from a 50-50 split other than to state that such a departure would occur frequently "by chance." However, when we obtained 9 heads, we had an uncomfortable feeling concerning the honesty of the coin. Nine heads out of 10 tosses is such a rare

occurrence that we begin to suspect that the explanation may be found in terms of the characteristics of the coin rather than in the so-called "laws of chance." The critical question is, "Where do we draw the line which determines what inferences we make about the coin?"

The answer to this question reveals the basic nature of science: its probabilistic rather than its absolutistic orientation. In the social sciences, most researchers have adopted one of the following two cutoff points as the basis for *inferring the operation of nonchance factors.*

1. When the event would occur *five percent* of the time or less, *by chance*, some researchers are willing to assert that the results are due to nonchance factors. This cutoff point is known variously as the *0.05 significance level*, or the *5.00% significance level.*
2. When the event would occur *one percent* of the time or less, *by chance*, other researchers are willing to assert that the results are due to nonchance factors. This cutoff point is known as the *0.01 significance level*, or the *1.00% significance level.*

The level of significance set by the experimenter for inferring the operation of nonchance factors is known as the *alpha* (α) *level*. Thus when employing the 0.05 level of significance, $\alpha = 0.05$; when employing the 0.01 level of significance, $\alpha = 0.01$.

In order to determine whether the results were due to nonchance factors in the present coin experiment, we need to calculate the probability of obtaining an event as *rare* as 9 heads out of 10 tosses. In determining the rarity of an event, we must consider the fact that the rare event can occur in both directions and that it includes more extreme events. In other words, the probability of an event as *rare* as 9 heads out of 10 tosses is equal to

$$p(9 \text{ heads}) + p(10 \text{ heads}) + p(1 \text{ head}) + p(0 \text{ heads}).$$

Since this distribution is symmetrical,

$$p(9 \text{ heads}) = p(1 \text{ head}) \qquad \text{and} \qquad p(10 \text{ heads}) = p(0 \text{ heads}).$$

Thus

$$p(9 \text{ heads}) + p(10 \text{ heads}) + p(1 \text{ head}) + p(0 \text{ heads}) = 2[p(9 \text{ heads}) + p(10 \text{ heads})].$$

These p-values may be obtained from Fig. 12.3 as follows:

$$p(9 \text{ heads}) = 0.010, \qquad \text{and} \qquad p(10 \text{ heads}) = 0.001.$$

Therefore the two-tailed probability of an event as rare as 9 heads out of 10

tosses is

$$2(0.010 + 0.001) = 0.022 \text{ or } 2.2\%.$$

Employing the 0.05 significance level ($\alpha = 0.05$), we would conclude that the coin was biased (i.e., the results were due to nonchance factors). However, if we employed the 0.01 significance level ($\alpha = 0.01$), we would not be able to assert that these results were due to nonchance factors.

11.5 TESTING STATISTICAL HYPOTHESES: NULL HYPOTHESIS AND ALTERNATIVE HYPOTHESIS

At this point, many students become disillusioned by the arbitrary nature of decision making in science. Let us examine the logic of statistical inference a bit further and see if we can resolve some of the doubts. Prior to the beginning of any experiment, the researcher sets up two mutually exclusive hypotheses: (1) The *null hypothesis* (H_0) which specifies hypothesized values for one or more of the *population parameters*. (2) The *alternative hypothesis* (H_1) which asserts that the *population parameter* is some value other than the one hypothesized. In the present coin experiment, these two hypotheses read as follows:

H_0: the coin is unbiased, that is, $P = Q = \frac{1}{2}$.
H_1: the coin is biased, that is, $P \neq Q \neq \frac{1}{2}$.

The alternative hypothesis may be either *directional* or *nondirectional*. When H_1 *only* asserts that the population parameter is different from the one hypothesized, it is referred to as a *nondirectional* or *two-tailed* hypothesis (for example, $P \neq Q \neq \frac{1}{2}$). Occasionally, H_1 is *directional* or *one-tailed*. In this instance, in addition to asserting that the population parameter is different from the one hypothesized, we assert the direction of that difference (for example, $P > Q$ or $P < Q$). In evaluating the outcome of an experiment, one-tailed probability values should be employed whenever our alternative hypothesis is directional.

11.5.1 The Notion of Indirect Proof

Careful analysis of the logic of statistical inference reveals that the null hypothesis can never be proved. For example, if we had obtained exactly 5 heads on 10 tosses of a coin, would this prove that the coin was unbiased? The answer is a categorical "No!" A bias, if it existed, might be of such a small magnitude that we failed to detect it in 10 trials. But what if we tossed the coin 100 times and obtained 50 heads? Wouldn't this prove something? Again, the same considerations apply. No matter how many times we tossed the coin, we could never exhaust the population of possible outcomes. We can make the assertion, however, that no basis exists for rejecting the hypothesis that the coin is biased.

How, then, can we prove the alternative hypothesis that the coin is biased? Again, we cannot prove the alternative hypothesis directly. Think, for the moment, of the logic involved in the following problem.

Draw two lines on a paper and determine whether they are of different lengths. You compare them and say, "Well, certainly they are not equal. Therefore they must be of different lengths." By rejecting equality (in this case, the null hypothesis) you assert that there is a difference.

Statistical logic operates in exactly the same way. We cannot prove the null hypothesis, nor can we directly prove the alternative hypothesis. However, if we can *reject* the null hypothesis, we can assert its alternative, namely, that the population parameter is some value other than the one hypothesized. Applied to the coin problem, if we can reject the null hypothesis that $P = Q = \frac{1}{2}$, we can assert the alternative, namely, that $P \neq Q \neq \frac{1}{2}$. Note that the proof of the alternative hypothesis is always *indirect*. We have proved it by rejecting the null hypothesis. On the other hand, since the alternative hypothesis can neither be proved nor disproved directly, we can *never prove the null hypothesis* by rejecting the alternative hypothesis. The strongest statement we are entitled to make in this respect is that we *failed to reject the null hypothesis*.

What, then, are the conditions for rejecting the null hypothesis? Simply this: when employing the 0.05 level of significance, you reject the null hypothesis when a given result occurs, by chance, 5% of the time or less. When employing the 0.01 level of significance, you reject the null hypothesis when a given result occurs, by chance, 1% of the time or less. Under these circumstances, of course, you affirm the alternative hypothesis.

In other words, one rejects the null hypothesis when the results occur, by chance, 5% of the time or less (or 1% of the time or less), *assuming that the null hypothesis is the true distribution*. That is, one assumes that the null hypothesis is true, calculates the probability on the basis of this assumption, and if the probability is small, one rejects the assumption.

For reasons stated above, R. A. Fisher, the eminent British statistician, has affirmed:

"In relation to any experiment we may speak of this hypothesis as the 'null hypothesis,' and it should be noted that the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. *Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis.*"* (Italics added).

11.6 TESTING STATISTICAL HYPOTHESES: THE TWO TYPES OF ERROR

You may now ask, "But aren't we taking a chance that we will be wrong in rejecting the null hypothesis? Is it not possible that we have, in fact, obtained a statistically rare occurrence by chance?"

* Fisher, R. A., *The Design of Experiments*. Edinburgh: Oliver & Boyd, 1935, p. 16.

The answer to this question must be a simple and humble, "Yes." This is precisely what we mean when we say that science is probabilistic. If there is any absolute statement that scientists are entitled to make, it is that we can never assert with complete confidence that our findings or propositions are true. There are countless examples in science in which an apparently firmly established conclusion has had to be modified in the light of further evidence.

In the coin experiment, even if all the tosses had resulted in heads, it is possible that the coin was not, in fact, biased. By chance, once in every 1024 experiments, "on the average," the coin will turn up heads 10 out of 10 times. When we employ the 0.05 level of significance, approximately 5% of the time we will be wrong when we reject the null hypothesis and assert its alternative.

These are some of the basic facts of the reality of inductive reasoning to which the student must adjust. The student of behavior who insists upon absolute certainty before he speaks on an issue is a student who has been mute throughout his years, and who will remain so the rest of his life (probably).

The above considerations have led statisticians to formulate two types of errors that may be made in statistical inference.

11.6.1 Type I Error (Type α Error)

In a type I error, we reject the null hypothesis when it is actually true. The probability of making a type I error is α. We have already pointed out that if we set our rejection point at the 0.05 level of significance, we will mistakenly reject H_0 approximately 5% of the time. It would seem, then, that in order to avoid this type of error, we should set the rejection level as low as possible. For example, if we were to set $\alpha = 0.001$, we would risk a type I error only about one time in every thousand. However, the lower we set α, the greater is the likelihood that we will make a type II error.

11.6.2 Type II Error (Type β Error)

In a type II error, we fail to reject the null hypothesis when it is actually false. *Beta* (β) is the probability of making a type II error. This type of error is far more common than a type I error. For example, if we accept the 0.01 level of significance as the basis of rejecting the null hypothesis and then conduct an experiment in which the result we obtained would have occurred, by chance, only 2% of the time, we cannot reject the null hypothesis. Consequently, we cannot claim an experimental effect even though there may very well be one.

It is clear, then, that the lower we set the rejection level, the less is the likelihood of a type I error, and the greater is the likelihood of a type II error. Conversely, the higher we set the rejection level, the greater the likelihood of a type I error, and the smaller the likelihood of a type II error.

The fact that the rejection level is set as low as it is attests to the conservatism of scientists, i.e., the greater willingness on the part of the scientist to make

an error in the direction of *failing* to claim a result than to make an error in the direction of *claiming* a result when he is wrong.

You may now ask, "How can we tell when we are making a type I or a type II error?" The answer is simple, "We can't." If we examine, once again, the logic of statistical inference, we shall see why. We have already stated that, with rare exceptions, we cannot or will not know the true values of a population. Without this knowledge, how can we know whether our sample statistics have approximated or have failed to approximate the true value? How can we know whether or not we have mistakenly rejected a null hypothesis? On the other hand, if we did know a population value, we could know whether or not we made an error. However, under these circumstances, the whole purpose of sampling statistics is vitiated. We collect samples and draw inferences from samples only because our population values are unknowable, for one reason or another. When they become known, the need for statistical inference is lost.

Is there no way, then, to know which experiments reporting significant results are accurate and which are not? The answer is a conditional "Yes." If we were to repeat the experiment and obtain similar results, we would have increased confidence that we were not making a type I error. For example, if we tossed our coin in a second series of 10 trials and obtained 9 heads, we would feel far more confident that our coin was biased. Parenthetically, repetition of experiments is one of the areas in which research in the social sciences is weakest. The general attitude is that a study is not much good unless it is "different" and is therefore making a novel contribution. Replicating experiments, when performed, frequently go unpublished. In consequence, we may feel assured that in studies employing the 0.05 significance level, approximately one out of every 20 studies which rejects the null hypothesis is making a type I error.*

CHAPTER SUMMARY

We have seen that one of the basic problems of inferential statistics involves estimating population parameters from sample statistics.

In inferential statistics, we are frequently called upon to compare our *obtained* values with *expected* values. The expected values are given by the appropriate sampling distribution, which is a theoretical probability distribution of the possible values of a sample statistic.

---

* The proportion is probably even higher since our methods of accepting research reports for publication are heavily weighted in terms of the statistical significance of the results. Thus if four identical studies were conducted independently and only one obtained results which permitted rejection of the null hypothesis, this one would most likely be published. There is virtually no way for the general scientific public to know about the three studies which failed to reject the null hypothesis.

We have seen how to construct sampling distributions. In the present chapter, we employed the binomial distribution to construct sampling distributions for discrete, two-category variables.

We have seen that there are two mutually exclusive and exhaustive statistical hypotheses in every experiment: the null hypothesis (H_0) and the alternative hypothesis (H_1).

If the outcome of an experiment is rare (here "rare" is defined as some arbitrary but accepted probability value), we reject the null hypothesis and assert its alternative. If the event is not rare (i.e., the probability value is *greater* than what we have agreed upon as being significant), we fail to reject the null hypothesis. However, in no event are we permitted to claim that we have *proved* H_0.

The experimenter is faced with two types of errors in establishing a cutoff probability value which he will accept as significant.

Type I: rejecting the null hypothesis when it is true.

Type II: accepting the null hypothesis when it is false.

The basic conservatism of the scientist causes him to establish a low level of significance, causing him to make type II errors more commonly than type I errors.

Without replication of experiments we have no basis for knowing when a type I error has been made, and even with replication, we cannot claim knowledge of absolute truth.

Finally, and perhaps most important, we have seen that scientific knowledge is probabilistic and not absolute.

Terms to Remember:

Population
Sample
Sampling distribution
Binomial distribution
Coefficients of the binomial
0.05 Significance level
(5% Significance level)
0.01 Significance level
(1% Significance level)
Alpha (α) level
Null hypothesis (H_0)
Alternative hypothesis (H_1)
Directional or one-tailed hypothesis
Nondirectional or two-tailed hypothesis
Type I error (type α)
Type II error (type β)

EXERCISES

1. Explain, in your own words, the nature of drawing inferences in behavioral science. Be sure to specify the types of risks that are taken and the ways in which the researcher attempts to keep these risks within specifiable limits.

2. Give examples of experimental studies in which
   a) a type I error would be considered more serious than a type II error.
   b) a type II error would be considered more serious than a type I error.
3. After completing a study in experimental psychology, Nelson W. concluded, "I have proved that no difference exists between the two experimental conditions." Criticize his conclusion according to the logic of drawing inferences in science.
4. Explain what is meant by the following statement: "It can be said that the purpose of any experiment is to provide the occasion for rejecting the null hypothesis."
5. An experimental psychologist hypothesizes that drive affects running speed. Assume that he has set up a study to investigate the problem employing two different drive levels. Formulate H_0 and H_1.
6. In a ten-item true-false examination,
   a) what is the probability that an unprepared student will obtain all correct answers by chance?
   b) if eight correct answers constitute a passing grade, what is the probability that he will pass?
   c) what are the odds against his passing?
7. Identify H_0 and H_1 in the following:
   a) The population mean in intelligence is 100.
   b) The proportion of Democrats in Watanabe County is not equal to 0.50.
   c) The population mean in intelligence is not equal to 100.
   d) The proportion of Democrats in Watanabe County is equal to 0.50.
8. Suppose that you are a personnel manager responsible for recommending the promotion of an employee to a high-level executive position. What type of error would you be making if:
   a) The hypothesis that he is qualified (H_0) is erroneously accepted.
   b) The hypothesis that he is qualified is erroneously rejected.
   c) The hypothesis that he is qualified is correctly accepted.
   d) The hypothesis that he is qualified is correctly rejected.
9. Construct a binomial sampling distribution when $N = 7$, $P = Q = \frac{1}{2}$, and answer the following questions: What is the probability that:
   a) Six or more will be in the P category?
   b) Four or fewer will be in the P category.
   c) One or fewer *or* 6 or more will be in the P category.
   d) Two or more will be in the P category.
10. A stock-market analyst recommends the purchase or sale of stock by his client on the basis of hypotheses he has formulated about the future behavior of these stocks. What type of error is he making if he claims:
   a) H_0: The stock will remain stable. *Fact:* It goes up precipitously.
   b) H_0: The stock will remain stable. *Fact:* It falls abruptly.
   c) H_0: The stock will remain stable. *Fact:* It shows only minor fluctuation about a central value.

11. An investigator sets $\alpha = 0.01$ for rejection of H_0. He conducts a study in which he obtains a p-value of 0.02 and fails to reject H_0. *Discuss:* Is it more likely that he is accepting a true or a false H_0?
12. *Comment:* a student of psychology has collected a mass of data to test 100 different null hypotheses. On completion of the analysis he finds that 5 of the 100 comparisons yield p-values ≤ 0.05. He concludes: "Using $\alpha = 0.05$, I have found a true difference in five of the comparisons."
13. *Comment:* An investigator has tested 500 different individuals for evidence of extrasensory perception (ESP). Employing $\alpha = 0.01$, he concludes, "I have found 6 individuals who demonstrated ESP."
14. Does the null hypothesis in a one-tailed test differ from the null hypothesis in a two-tailed test?
15. Does the alternative hypothesis in a one-tailed test differ from the alternative hypothesis in a two-tailed test? Give an example.
16. In rejecting the null hypothesis for a one-tailed test, do all deviations count equally, Explain.
17. Discuss the similarities and differences between the normal curve and the binomial curve.

Statistical Inference and Continuous Variables

12

12.1 INTRODUCTION

In Chapter 11 we saw that one of the basic problems of inferential statistics is estimating population parameters from sample statistics. In doing so, we compare the values obtained from a sample with "expected" values which are given by the appropriate sampling distribution. Since the sampling distribution is a theoretical *probability* distribution, the crux of the problem becomes: *How do we determine these probability values?*

In Chapter 11 we employed the binomial distribution to determine the probability values for discrete two-category variables. In this chapter we will be dealing with continuously distributed variables. We previously pointed out (Section 10.6) that, with continuous variables, probability is defined in terms of areas under a curve. Let us now investigate the application of the concepts of inferential statistics to continuously distributed variables by looking at the construction of a sampling distribution of means.

12.2 SAMPLING DISTRIBUTION OF THE MEAN

In Table O we provide several different approximately normally distributed populations. Let us conduct a hypothetical sampling experiment with one of these populations which will serve to clarify many of the concepts we shall subsequently develop.

Imagine that we randomly draw (with replacement)* a sample of two cases from the population in which $\mu = 5.00$ and $\sigma = 0.99$. For example, we might draw scores of 3 and 6. We calculate the sample mean and find $\overline{X} = 4.5$. Now, suppose we continue to draw samples of $N = 2$ (e.g., we might draw scores of 2, 8; 3, 7; 4, 5; 6, 6; etc.) until we obtain an indefinitely large number of samples. If we calculate the sample mean for each sample drawn, and treat each of these

* If the population is infinite or extremely large, the difference between sampling with or without replacement is negligible.

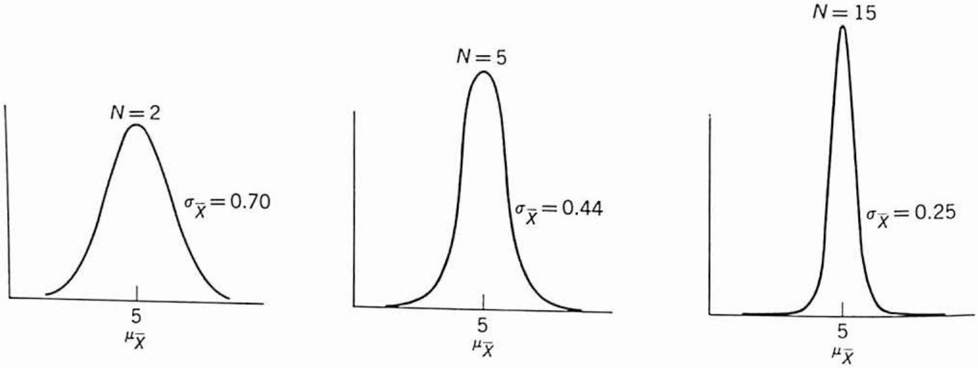


Fig. 12.1 Frequency curves of sample means drawn from a population in which $\mu = 5.00$ and $\sigma = 0.99$.

sample means as a raw score, we may set up a frequency distribution of these sample means.

Let us repeat the above procedures with increasingly larger sample sizes, for example, $N = 5$, $N = 15$. We now have three frequency distributions of sample means based on three different sample sizes.

Intuitively, what might we expect these distributions to look like? Since we are selecting at random from the population, we would expect the mean of the distribution of sample means to approximate the mean of the population.

How might the dispersion of these sample means compare with the variability in the original distribution of scores? In the original distribution, when $N = 1$, the probability of obtaining a score as extreme as say, 8 is $\frac{4}{1000}$ or 0.004 (see Table O). The associated probability of obtaining a sample mean equal to 8 when $N = 2$ (i.e., drawing scores of 8, 8) is equal to $\frac{4}{1000} \times \frac{4}{1000}$ or 0.000016 (formula 10.10). Clearly, when $N = 5$, the probability of obtaining results this extreme (for example, $\overline{X} = 8$) becomes exceedingly small. In other words, the probability of drawing extreme values of the sample mean is smaller as N increases. Since the standard deviation is a direct function of the number of extreme scores (see Chapter 6), it follows that a distribution containing proportionately fewer extreme scores will have a lower standard deviation. Therefore, if we treat each of the sample means as a raw score and then calculate the standard deviation ($\sigma_{\overline{X}}$ referred to as the *standard error of the mean**), it is clear that as N increases the variability of the sample means decreases.

If these sampling experiments were actually conducted, the above frequency polygons of sample means would be obtained (Fig. 12.1).

---

* This notation represents the standard deviation of a sampling distribution of means. This is purely a theoretical notation since, with an infinite number of sample means, it is not possible to assign a specific value to the number of sample means involved.

There are three important lessons which may be learned from a careful examination of Fig. 12.1.

1. The distribution of sample means, drawn from a normally distributed population, tends to be bell-shaped or "normal." Indeed, it can be shown that even if the underlying distribution is skewed, the distribution of sample means will tend to be normal.

2. The mean of the sample means ($\mu_{\overline{X}}$) is equal to the mean of the population (μ) from which these samples were drawn.

3. The distribution of sample means becomes more and more compact as we increase the size of the sample. This is an extremely important point in statistical inference, about which we shall have a great deal more to say shortly.

If we base our estimate of the population mean on a *single* sample drawn from the population, our approximation to the parameter is likely to be closer as we increase the size of the sample. In other words, if it is true that the dispersion of sample means decreases with increasing sample size, it also follows that the mean of any single sample is more likely to be closer to the mean of the population as the sample size increases

These three observations illustrate a rather startling theorem which is of fundamental importance in inferential statistics, i.e., the *central-limit theorem.*

The **central-limit theorem** states: If random samples of a fixed N are drawn from *any* population (regardless of the form of the population distribution), as N becomes larger, the distribution of sample means approaches normality with the overall mean approaching μ, the variance of the sample means $\sigma_{\overline{X}}^2$ being equal to σ^2/N, and a standard error $\sigma_{\overline{X}}$ of $\sigma/\sqrt{N}$.

Stated symbolically,

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{N}, \tag{12.1}$$

and

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}. \tag{12.2}$$

12.3 TESTING STATISTICAL HYPOTHESES: PARAMETERS KNOWN

Let us briefly examine some of the implications of the relationships we have just discussed.

When μ and σ are *known* for a given population, it is possible to describe the form of the distribution of sample means when N is large (regardless of the form of the original distribution). It will be a normal distribution with a mean ($\mu_{\overline{X}}$) equal to μ and a standard error ($\sigma_{\overline{X}}$) equal to $\sigma/\sqrt{N}$. It now becomes possible to determine probability values in terms of areas under the normal curve. Thus we may use the known relationships of the normal probability curve to determine the probabilities associated with any sample mean (of a given N) randomly drawn from this population.

We have already seen (Section 7.3) that any normally distributed variable may be transformed into the normally distributed z-scale. We have also seen (Section 10.8) that we may establish probability values in terms of the relationships between z-scores and areas under the normal curve. That is, for any given raw score value (X) with a certain proportion of area beyond it, there is a corresponding value of z with the same proportion of area beyond it. Similarly, for any given value of a sample mean ($\overline{X}$) with a certain proportion of area beyond it, there is a corresponding value of z with the same proportion of area beyond it. Thus assuming that the form of the distribution of sample means is normal, we may establish probability values in terms of the relationships between z-scores and areas under the normal curve.

To illustrate: Given a population with $\mu = 250$ and $\sigma = 50$ from which we randomly select 100 scores ($N = 100$), what is the probability that the sample mean ($\overline{X}$) will be equal to or greater than 255? Thus $H_0: \mu = \mu_0 = 250$.

The value of z corresponding to $\overline{X} = 255$ is obtained as follows:

$$z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}, \tag{12.3}$$

where $\mu_0 =$ value of the population mean under H_0,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{50}{\sqrt{100}} = 5.00,$$

and

$$z = \frac{255 - 250}{5.00} = 1.00.$$

Looking up a z of 1.00 in Column C (Table A), we find that 15.87% of the sample means fall at or above $\overline{X} = 255$. Therefore there are approximately 16 chances in 100 of obtaining a sample mean equal to or greater than 255 from this population when $N = 100$.

By a simple extension of the above logic, we may entertain hypotheses concerning the values of parameters of the population from which the sample mean was drawn. For example, given a sample mean of 263 for $N = 100$, is it reasonable to assume that this sample was drawn from the above population?

Let us set up this problem in more formal statistical terms:

1. *Null hypothesis* (H_0): The mean of the population (μ) from which the sample was drawn equals 250, that is, $\mu = \mu_0 = 250$.

2. *Alternative hypothesis* (H_1): The mean of the population from which the sample was drawn does *not* equal 250; $\mu \neq \mu_0$. Note that H_1 is nondirectional; consequently, a two-tailed test of significance will be employed.

3. *Significance level:* $\alpha = 0.01$. If the difference between the sample mean and the specified population mean is so extreme that its associated probability of occurrence under H_0 is equal to or less than 0.01, we will reject H_0.

4. *Critical region for rejection of* H_0: $|z| \geq |z_{0.01}| = 2.58$*. *A critical region is that portion of area under the curve which includes those values of a statistic which lead to rejection of the null hypothesis.*

The critical region is chosen to correspond with the selected level of significance. Thus for $\alpha = 0.01$, two-tailed test, the critical region is bounded by those values of $z_{0.01}$ which mark off a total of 1% of the area. Referring to Column C (Table A), we find that the area beyond a z of 2.58 is approximately 0.005. We double 0.005 to account for both tails of the distribution. Figure 12.2 depicts the critical region for rejection of H_0 when $\alpha = 0.01$, two-tailed test.

Therefore, in order to reject H_0 at the 0.01 level of significance, the absolute value of the obtained z must be equal to or greater than $|z_{0.01}|$ or 2.58. Similarly, to reject H_0 at the 0.05 level of significance, the absolute value of the obtained z must be equal to or greater than $|z_{0.05}|$ or 1.96.

In the present example, the value of z corresponding to $\overline{X} = 263$ is

$$z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{263 - 250}{5.00} = 2.60.$$

Decision: Since the obtained z falls within the critical region, (that is, $2.60 > z_{0.01}$), we may reject H_0 at the 0.01 level of significance.

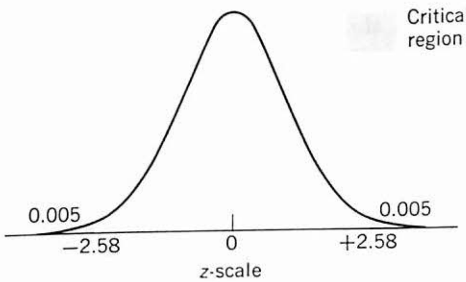


Fig. 12.2 Critical region for rejection of H_0 when $\alpha = 0.01$, two-tailed test.

12.4 ESTIMATION OF PARAMETERS: POINT ESTIMATION

So far, we have been concerned with testing hypotheses when the population parameters are known. However, we have taken some pains in this book to point out that population values are rarely known. The fact that we do not know the population values does not prevent us from using the above logic.

Whenever we make inferences about population parameters from sample data, we compare our sample results with the expected results given by the appropriate sampling distribution. A hypothetical sampling distribution of sample means is associated with any sample mean. This distribution has a mean, $\mu_{\overline{X}}$, and a standard deviation, $\sigma_{\overline{X}}$. So far, in order to obtain the values of

---

* Since $z_{0.01} = \pm 2.58$, $|z| \geq |z_{0.01}|$ is equivalent to stating $z \geq 2.58$ or $z \leq -2.58$.

$\mu_{\overline{X}}$ and $\sigma_{\overline{X}}$, we have required a knowledge of μ and σ. In the absence of knowledge concerning the exact values of the parameters, we are forced to estimate μ and σ from the statistics calculated from sample data. Since, in actual practice, we rarely select more than one sample, our estimates are generally based on the statistics calculated from a single sample. All such estimates, involving the use of single sample values, are known as *point estimates.*

12.4.1 Estimating $\sigma_{\overline{X}}$ from Sample Data

You will recall that we previously defined the variance of a sample as $s^2 = \sum(X - \overline{X})^2/N$ in formula (6.2). We obtained the standard deviation, s, by finding the square root of this value. These definitions are perfectly appropriate as long as we are interested only in describing the variability of a sample. However, when our interest shifts to estimating the population variance from a sample value we find that the above definition is inadequate since $\sum(X - \overline{X})^2/N$ tends, on the average, to *underestimate* the population variance. In other words, it provides a *biased estimate* of the population variance whereas an unbiased estimate is required.

We shall define an *unbiased estimate* as an estimate which equals, on the average, the value of the parameter. That is, when we make the statement that a statistic is an unbiased estimate of a parameter, we are saying that the mean of the distribution of an extremely large number of sample statistics, drawn from a given population, tends to center upon the corresponding value of the parameter. It has been demonstrated that an unbiased estimate of the population variance may be obtained by dividing the sum of squares by $N - 1$. We shall employ the symbol $\hat{s}^2$ to represent a sample variance providing an *unbiased estimate of the population variance*, and $\hat{s}$ to represent a sample standard deviation based on the unbiased variance estimate. Thus

$$\text{unbiased estimate of } \sigma^2 = \hat{s}^2 = \frac{\sum(X - \overline{X})^2}{N - 1}, \tag{12.4}$$

and

$$\text{estimated } \sigma = \hat{s} = \sqrt{\hat{s}^2}. \tag{12.5}$$

We are now able to estimate $\sigma^2_{\overline{X}}$ and $\sigma_{\overline{X}}$ from sample data. We shall employ the symbols $s^2_{\overline{X}}$ and $s_{\overline{X}}$ to refer to the estimated variance and standard error of the mean, respectively. Since we do not know σ^2, we accept the unbiased variance estimate ($\hat{s}^2$) as the best estimate we have of the population variance. Thus the formula for determining the variance of the mean from sample data is

$$\text{estimated } \sigma^2_{\overline{X}} = s^2_{\overline{X}} = \hat{s}^2/N. \tag{12.6}$$

We estimate the standard error of the mean by finding the square root of this value:

$$\text{estimated } \sigma_{\overline{X}} = s_{\overline{X}} = \sqrt{\hat{s}^2/N} = \hat{s}/\sqrt{N}. \tag{12.7}$$

If the sample variance (not the unbiased estimate) is used, we may estimate $\sigma_{\overline{X}}$ as

$$\text{estimated } \sigma_{\overline{X}} = s_{\overline{X}} = \frac{s}{\sqrt{N-1}} = \sqrt{\frac{\sum(X-\overline{X})^2}{N(N-1)}}. \tag{12.8}*$$

Formula (12.8) is the one most frequently employed in the behavioral sciences to estimate the standard error of the mean. We shall follow this practice.

Before proceeding further, let us briefly review some of the symbols we have been discussing:

μ	Population mean
μ_0	Value of the population mean under H_0
$\overline{X}$	Sample mean
σ^2	Population variance
s^2	Sample variance, $\sum(X-\overline{X})^2/N$
$\hat{s}^2$	Unbiased estimate of population variance, $\sum(X-\overline{X})^2/(N-1)$
σ	Population standard deviation
s	Sample standard deviation
$\hat{s}$	Sample standard deviation based upon the unbiased variance estimate
$\mu_{\overline{X}}$	Mean of the sampling distribution of sample means
$\sigma_{\overline{X}}$	Standard error of the mean (theoretical), $\sigma/\sqrt{N}$
$s_{\overline{X}}$	Estimated standard error of the mean, $\hat{s}/\sqrt{N}$ or $s/\sqrt{N-1}$

* The following algebraic proof demonstrates how we arrive at this estimate of $\sigma_{\overline{X}}$:

$$s^2 = \frac{\sum(X-\overline{X})^2}{N},$$

multiplying both sides of the equation by $N/(N-1)$,

$$\frac{N}{N-1}s^2 = \frac{N\sum(X-\overline{X})^2}{N(N-1)} = \hat{s}^2 \quad \text{[see formula (12.4)].}$$

Thus from formula (12.7),

$$s_{\overline{X}} = \sqrt{\frac{\hat{s}^2}{N}} = \sqrt{\frac{Ns^2}{N(N-1)}} = \frac{s}{\sqrt{N-1}}.$$

12.5 TESTING STATISTICAL HYPOTHESES WITH UNKNOWN PARAMETERS: STUDENT'S *t*

We previously pointed out that when the parameters of a population are known, it is possible to describe the form of the sampling distribution of sample means. It will be a normal distribution with $\sigma_{\overline{X}}$ equal to $\sigma/\sqrt{N}$. By employing the relationship between the z-scale and the normal distribution, we were able to test hypotheses using $z = (\overline{X} - \mu_0)/\sigma_{\overline{X}}$ as a test statistic. When σ is not known, we are forced to estimate its value from sample data. Consequently, estimated $\sigma_{\overline{X}}$ (that is, $s_{\overline{X}}$) must be based on the estimated σ (that is, $\hat{s}$), that is, $s_{\overline{X}} = \hat{s}/\sqrt{N}$. Now, if substituting $\hat{s}$ for σ provided a reasonably good approximation to the sampling distribution of means, we could continue to use z as our test statistic, and the normal curve as the model for our sampling distribution. As a matter of fact, however, this is not the case. At the turn of the century, a statistician by the name of William Gosset, who published under the pseudonym of Student, noted that the approximation of $\hat{s}$ to σ is poor, particularly for small samples. This failure of approximation is due to the fact that, with small samples, $\hat{s}$ will tend to underestimate σ more than one half of the time. Consequently, the statistic

$$\frac{\overline{X} - \mu_0}{\hat{s}/\sqrt{N}}$$

will tend to be spread out more than the normal distribution.

Gosset's major contribution to statistics consisted of his description of a distribution, or rather, a family of distributions, which permits the testing of hypotheses with samples drawn from normally distributed populations, when σ is not known. These distributions are referred to variously as the **t-distributions** or **Student's t**. The ratio employed in the testing of hypotheses is known as the t-ratio:

$$t = \frac{\overline{X} - \mu_0}{s_{\overline{X}}}, \tag{12.9}$$

where μ_0 is the value of the population mean under H_0.

The t-statistic is similar in many respects to the previously discussed z-statistic. Both statistics are expressed as the deviation of a sample mean from a population mean (known or hypothesized) in terms of the standard error of the mean. By reference to the appropriate sampling distribution, we may express this deviation in terms of probability. When the z-statistic is used, the standard normal curve is the appropriate sampling distribution. For the t-statistic there is a family of distributions which vary as a function of *degrees of freedom* (df).

The term "degrees of freedom" refers to the number of values which are free to vary after we have placed certain restrictions on our data. To illustrate, if

we had four numbers on which we placed the restriction that the sum must equal 115, it is clear that three numbers could take on any value (i.e., are free to vary) whereas the fourth would be fixed. Thus, if three values are 30, 45, and 25, respectively, the fourth must be 15 in order to make the sum equal to 115. The number of degrees of freedom, then, is $N - 1$, or 3. Generalizing, for any given sample on which we have placed a single restriction, the number of degrees of freedom is $N - 1$.

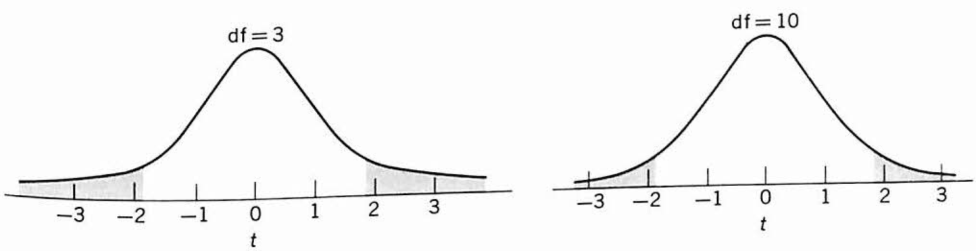
Note that when $s/\sqrt{N-1}$ [formula (12.8)] is employed to obtain $s_{\overline{X}}$, the quantity under the square root sign $(N - 1)$ is the degrees of freedom.

12.5.1 Characteristics of *t*-Distributions

Let us compare the characteristics of the *t*-distributions with the already familiar standard normal curve. First, both distributions are symmetrical about a mean of zero. Therefore the proportion of area beyond a particular positive *t*-value is equal to the proportion of area below the corresponding negative *t*.

Secondly, the *t*-distributions are more spread out than the normal curve. Consequently, the proportion of area beyond a specific value of *t* is *greater* than the proportion of area beyond the corresponding value of *z*. However, the greater the df, the more the *t*-distributions resemble the standard normal curve. In order that you may see the contrast between the *t*-distributions and the normal curve we have reproduced three curves in Fig. 12.3: the sampling distributions of *t* when df = 3, df = 10, and the normal curve.

Inspection of Fig. 12.3 permits several interesting observations. We have already seen that with the standard normal curve, $|z| \geq 1.96$ defines the region of rejection at the 0.05 level of significance. However, when df = 3, a $|t| \geq 1.96$ includes approximately 15% of the total area. Consequently, if we were



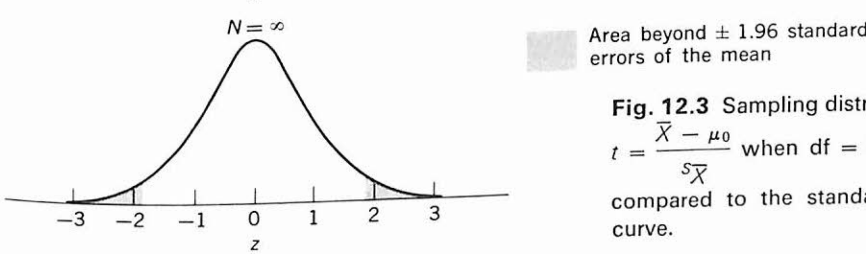


Fig. 12.3 Sampling distributions of $t = \dfrac{\overline{X} - \mu_0}{s_{\overline{X}}}$ when df = 3, and 10, compared to the standard normal curve.

to employ the normal curve for testing hypotheses when N is small (therefore df is small) and σ is unknown, we would be in serious danger of making a type I error, i.e., rejecting H_0 when it is true. Obviously, a much larger value of t is required to mark off the bounds of the critical region of rejection. Indeed, when df $= 3$, the absolute value of the obtained t must be equal to or greater than 3.18 to reject H_0 at the 0.05 level of significance (two-tailed test). However, as df increases, the differences in the proportions of area under the normal curve and the Student t-distributions become negligible.

In contrast to our use of the normal curve, the tabled values for t (Table C) are *critical values*, i.e., *those values which bound the critical rejection regions corresponding to varying levels of significance.* Thus in using the table for the distributions of t, we locate the appropriate number of degrees of freedom in the left-hand column, and then find the column corresponding to the chosen α. The tabled values represent the t-ratio required for significance. If the absolute value of our obtained t-ratio equals or exceeds this tabled value, we may reject H_0.

12.5.2 Illustrative Problem: Student's *t*

Let us now examine a hypothetical case involving a small sample.

A group of 17 ninth grade students selected on the basis of an expressed "interest" in science were given a test measuring their knowledge of basic scientific concepts. The list was constructed to yield a normal distribution with a mean for ninth graders equal to 78. The results of the study were:

$$\overline{X} = 84, \qquad s = 16, \qquad N = 17,$$

Is it reasonable to assume that these ninth graders are representative of a population in which $\mu = 78$?

Let us set up this problem in more formal statistical terms.

1. *Null hypothesis* (H_0): The mean of the population from which the sample was drawn equals 78 ($\mu = \mu_0 = 78$).
2. *Alternative hypothesis* (H_1): The mean of the population from which the sample was drawn does *not* equal 78 ($\mu \neq \mu_0$).
3. *Statistical test:* The Student t-ratio is chosen because we are dealing with a normally distributed variable in which σ is unknown.
4. *Significance level:* $\alpha = 0.05$.
5. *Sampling distribution:* The sampling distribution is the Student t-distribution with df $= 16$.
6. *Critical region:* $t_{0.05} \geq 2.12$. Since H_1 is nondirectional, the critical region consists of all values of $t \geq 2.12$ and $t \leq -2.12$.

In the present example, the value of t corresponding to $\overline{X} = 84$ is

$$t = \frac{\overline{X} - \mu_0}{s_{\overline{X}}} = \frac{84 - 78}{16/\sqrt{16}} = 1.50.$$

Decision: Since the obtained t does not fall within the critical region (that is, $1.50 < t_{0.05}$), we accept H_0.

12.6 ESTIMATION OF PARAMETERS: INTERVAL ESTIMATION

We have repeatedly pointed out that one of the basic problems in inferential statistics is the estimation of the parameters of a population from statistics calculated from a sample. This problem, in turn, involves two subproblems: (1) *point estimation*, and (2) *interval estimation*.

When we estimate parameters employing single sample values, these estimates are known as *point estimates*. A single sample value drawn from a population provides an estimate of the population parameter. But, how good an estimate is it? If a population mean were known to be 100, would a sample mean of 60 constitute a good estimate? How about a sample mean of 130, 105, 98, or 99.4? Under what conditions do we consider an estimate good? Since we know that the population parameters are virtually never known and that we generally employ samples to estimate these parameters, is there any way to determine the amount of error we are likely to make? The answer to this question is a negative one. However, it is possible, not only to estimate the population parameter (point estimation), but also to state a range of values within which we are confident that the parameter falls (*interval estimation*). Moreover, we may express our confidence in terms of probability theory.

Let us say that we are to estimate the weight of a man based on physical inspection. Let us assume that we are unable to place him on a scale, and that we cannot ask him his weight. This problem is similar to many we have faced throughout this text. We cannot know the population value (the man's true weight) and hence we are forced to estimate it. Let us say that we have the impression that he weighs about 200 pounds. If we are asked, "How confident are you that he weighs *exactly* 200 pounds?," we would probably reply, "I doubt that he weighs exactly 200 pounds. If he does, you can credit me with a fantastically lucky guess. However, I feel reasonably confident that he weighs between 190 and 210 pounds." In doing this, we have stated the interval within which we feel confident that the true weight falls. After a moment's reflection, we might hedge slightly, "Well, he is almost certainly between 180 and 220 pounds. In any event, I feel perfectly confident that his true weight falls somewhere between 170 and 230 pounds." Note that the greater the size of the interval, the greater is our feeling of certitude that the true value is encompassed

between these limits. Note also that in stating these confidence limits, we are, in effect, making two statements: (1) We are stating the limits between which we feel our subject's true weight falls, and (2) we are rejecting the possibility that his true weight falls outside of these limits. Thus, if someone asks, "Is it conceivable that our subject weighs as much as 240 pounds or as little as 160 pounds?," our reply would be a negative one.

12.7 CONFIDENCE INTERVALS AND CONFIDENCE LIMITS

In our preceding example, we were, in a sense, concerning ourselves with the problem of estimating *confidence limits.* In effect, we were attempting to determine the interval within which any hypotheses concerning the weight of the man might be considered tenable and outside which any hypotheses would be considered untenable. The interval within which we consider hypotheses tenable is known as the *confidence interval,* and the limits defining the interval are referred to as *confidence limits.*

Let us look at a sample problem and apply our statistical concepts to the estimation of confidence intervals.

A school district is trying to decide on the feasibility of setting up a vocational training program in its public high school curriculum. In part, the decision will depend on estimates of the average I.Q. of high school students within the district. With only one school psychologist in the district, it is impossible to administer an individual test to each student. Consequently, we must content ourselves with testing a random sample of students and base our estimates on this sample. We administer this test to a random sample of 26 students and obtain the following results:

$$\overline{X} = 108,$$
$$s = 15,$$
$$N = 26.$$

Our best estimate of the population mean (i.e., the mean I.Q. of children within the school district) is 108. However, even though our sample statistics provide our best estimates of population values, we recognize that such estimates are subject to error. As with the weight problem, we would be fantastically lucky if the mean I.Q. of the high school population were actually 108. On the other hand, if we have employed truly random selection procedures, we have a right to believe that our sample value is fairly close to the population mean. The critical question becomes: Between what limits will we entertain, as tenable, hypotheses concerning the value of the population mean (μ) in intelligence?

We have seen that the mean of the sampling distribution of sample means ($\mu_{\overline{X}}$) is equal to the mean of the population. We have also seen that, since, for

any given N, we may determine how far sample means are likely to deviate from any given or hypothesized value of μ, we may determine the likelihood that a particular $\overline{X}$ could have been drawn from a population with a mean of μ_0, where μ_0 represents the value of the population mean under H_0. Now, since we do not know the value of the population mean, we are free to hypothesize *any* value we desire.

It should be clear that we could entertain an unlimited number of hypotheses concerning the population mean and subsequently reject them, or fail to reject them, on the basis of the size of the t-ratios. For example, in the present problem, let us select a number of hypothetical population means. We may employ the 0.05 level of significance (two-tailed test), and test the hypothesis that $\mu_0 = 98$. The value of t corresponding to $\overline{X} = 108$ is

$$t = \frac{108 - 98}{15/\sqrt{25}} = 3.333.$$

In Table C, we find that $t_{0.05}$ for 25 df is 2.060. Since our obtained t is greater than this critical value, we reject the hypothesis that $\mu_0 = 98$. In other words, it is unlikely that $\overline{X} = 108$ was drawn from a population with a mean of 98.

Our next hypothesis is that $\mu_0 = 100$, which gives a t of

$$t = \frac{108 - 100}{3} = 2.667.$$

Since $2.667 > t_{0.05}$ (or 2.060), we may reject the hypothesis that the population mean is 100.

If we hypothesize $\mu_0 = 102$, the resulting t-ratio of 2.000 is less than $t_{0.05}$. Consequently, we may consider the hypothesis that $\mu_0 = 102$ tenable. Similarly, if we obtained the appropriate t-ratios, we would find that the hypothesis $\mu_0 = 114$ is tenable, whereas hypotheses of values greater than 114 are untenable. Thus $\overline{X} = 108$ was probably drawn from a population whose mean falls in the interval 102–114 (note that these limits, 102 and 114, represent approximate limits, i.e., the closest integers). The hypothesis that $\overline{X} = 108$ was drawn from a population with $\mu < 102$ or $\mu > 114$ may be rejected at the 0.05 level of significance. The interval within which the population mean probably lies is called the *confidence interval*. We refer to the limits of this interval as the *confidence limits*. Since we have been employing $\alpha = 0.05$, we call it the *95 percent confidence interval*. Similarly, if we employed $\alpha = 0.01$, we could obtain the *99 percent confidence interval*.

It is not necessary to perform all the above calculations to establish the confidence limits. We may calculate the exact limits of the 95 percent confidence interval directly.

To determine the upper limit for the 95 percent confidence interval:

$$\text{upper limit } \mu_0 = \overline{X} + t_{0.05}\,(s_{\overline{X}}). \tag{12.10}$$

Similarly, for the lower limit,

$$\text{lower limit } \mu_0 = \overline{X} - t_{0.05}\,(s_{\overline{X}}). \tag{12.11}$$

For the 99 percent confidence interval, merely substitute $t_{0.01}$ in the above formulas.

You will note that these formulas are derived algebraically from formula (12.9):

$$t_{0.05} = \frac{\overline{X} - \mu_0}{s_{\overline{X}}}, \qquad \text{therefore } \mu_0 = \overline{X} + t_{0.05}(s_{\overline{X}}).$$

Employing formula (12.10), we find that the upper 95 percent confidence limit in the above problem is

$$\begin{aligned} \text{upper limit } \mu_0 &= 108 + (2.060)(3.0) \\ &= 108 + 6.18 = 114.18. \end{aligned}$$

Similarly, employing formula (13.11), we find that the lower confidence limit is

$$\begin{aligned} \text{lower limit } \mu_0 &= 108 - (2.060)(3.0) \\ &= 108 - 6.18 = 101.82. \end{aligned}$$

Having established the lower and the upper limits as 101.82 and 114.18, respectively, we may now conclude: On the basis of our obtained mean and standard deviation, which were computed from scores drawn from a population in which the true mean is unknown, we assert that the population mean probably falls within the interval which we have established. Since the probability that the population mean lies outside these limits is 5% ($\alpha = 0.05$), the probability that this interval contains the population mean is 95%. In other words, in the long run, we will be correct 95% of the time when we state that the true mean lies within the 95% confidence interval.

Some words of caution in interpreting the confidence interval. In establishing the interval, within which we believe the population mean falls, we have *not* established any probability that our obtained mean is correct. In other words, we cannot claim that the chances are 95 in 100 that the population mean is 108. Our statements are valid only with respect to the interval and not with respect to any particular value of the sample mean. In addition, since the population mean is a fixed value and does not have a distribution, our probability statements never refer to μ. The probability we assert is about the interval, i.e., the probability that the interval contains μ.

Finally, when we have established the confidence interval of the mean, we are not stating that the probability is 0.95 that the particular interval we have calculated contains the population mean. It should be clear that, if we were to select repeated samples from a population, both the sample means and the standard deviations would differ from sample to sample. Consequently, our estimates of the confidence interval would also vary from sample to sample. When we have established the 95% confidence interval of the mean, then, we are stating that, if repeated samples of a given size are drawn from the population, 95 percent of the interval estimates will include the population mean.

12.8 TEST OF SIGNIFICANCE FOR PEARSON r, ONE-SAMPLE CASE

In Chapter 8 we discussed the calculation of two statistics—the Pearson r and r_{rho}— commonly employed to describe the extent of the relationship between two variables. It will be recalled that the coefficient of correlation varies between ± 1.00, with $r = 0.00$ indicating the absence of a relationship. It is easy to overlook the fact that correlation coefficients based on sample data are only estimates of the corresponding population parameter and, as such, will distribute themselves about the population value. Thus it is quite possible that a sample drawn from a population in which the true correlation is zero may yield a high positive or negative correlation *by chance*. The null hypothesis most often investigated in the one-sample case is that the *population correlation coefficient* (ρ) is zero.

It is clear that a test of significance is called for. However, the test is complicated by the fact that the sampling distribution of ρ is usually non-normal, particularly as ρ approaches the limiting values of ± 1.00. Consider the case in which ρ equals $+0.80$. It is clear that sample correlation coefficients drawn from this population will distribute themselves around $+0.80$ and can take on any value from -1.00 to $+1.00$. It is equally clear, however, that there is a definite restriction in the range of values that sample statistics greater than $+0.80$ can assume whereas there is no similar restriction for values less than $+0.80$. The result is a negatively skewed sampling distribution. The departure from normality will, in general, be less as the number of paired scores in the sample increases. When the population correlation from which the sample is drawn is equal to zero, the sampling distribution is more likely to be normal. These relationships are demonstrated in Fig. 12.4, which illustrates the sampling distribution of the correlation coefficient when $\rho = -0.80$, 0, and $+0.80$.

Fisher has described a procedure for transforming sample r's to a statistic z_r, which yields a sampling distribution more closely approximating the normal curve, even for samples employing small n's. The transformation to z_r's involves the use of the following formula:

$$z_r = \tfrac{1}{2} \log_e (1 + r) - \tfrac{1}{2} \log_e (1 - r) \tag{12.12}$$

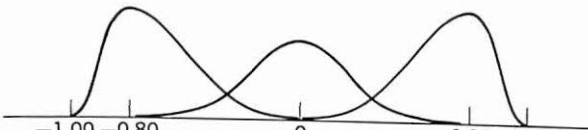


Fig. 12.4 Illustrative sampling distributions of correlation coefficients when $\rho = -0.80$, and $+0.80$.

The test statistic is the normal deviate, z, in which

$$z = \frac{z_r - Z_r}{\sqrt{\dfrac{1}{n-3}}}, \tag{12.13}$$

where

$z_r =$ the transformed value of the sample r, and

$Z_r =$ the transformed value of the population correlation coefficient specified under H_0.

Obtaining z_r is greatly simplified by Table F which shows the value of z_r, corresponding to each value of r, in steps of 0.01, between 0.00 and 0.99. Thus, referring to Table F, we see that an r of 0.60, for example, has a corresponding z_r of 0.693.

Let us look at an illustrative example: A social psychologist has developed a scale which purports to measure the degree of submission to authority. He correlates the scores made on the scale by 15 subjects with their scores on an inventory which reveals the degree of prejudice felt toward minority groups. He obtains a Pearson r of 0.60. May he conclude that the obtained correlation is not likely to have been drawn from a population in which the true correlation is zero?

Let us set up this problem in formal statistical terms.

1. *Null hypothesis* (H_0): The population correlation coefficient from which this sample was drawn equals 0.00 ($\rho = 0.00$).
2. *Alternative hypotheses* (H_1): The population correlation coefficient from which the sample was drawn does *not* equal 0.00 ($\rho \neq 0.00$).
3. *Statistical test:* The z-test based on Fisher's transformation of r and ρ to z_r and Z_r respectively.
4. *Significance level:* $\alpha = 0.05$, two-tailed test.
5. *Sampling distribution:* The normal probability curve.

6. *Critical region:* $|z_{0.05}| \geq |1.96$. Since H_1 is nondirectional, the critical region consists of all values of $z \geq 1.96$ and $z \leq -1.96$.

In the present example, the value of z_r corresponding to $r = 0.60$ is 0.693 and $Z_r = 0.00$. Thus

$$z = \frac{0.693 - 0.00}{\sqrt{\frac{1}{15 - 3}}} = \frac{0.693}{0.289} = 2.40.$$

Decision: Since the obtained $z > z_{0.05}$, it falls within the critical region for rejecting H_0. Thus it may be concluded that the sample was drawn from a population in which $\rho > 0.00$.

12.8.1 Test of Significance of r_{rho}, One-Sample Case

Table G presents the critical values of r_{rho}, one- and two-tailed tests, for selected values of n from 5 to 30.

In Section 8.5 we demonstrated the calculation of r_{rho} from data consisting of 15 pairs of ranked scores. A correlation of 0.64 was found.

Since $n = 15$ is not listed in Table G, it is necessary to interpolate, employing the critical values for $n = 14$ and 16. The critical value at the 0.05 level, two-tailed test, for $n = 14$ is 0.544; at $n = 16$, it is 0.506. By employing linear interpolation, we may roughly approximate the critical value corresponding to $n = 15$. Linear interpolation involves subtracting the latter from the former, dividing by $\frac{1}{2}$ (since $n = 15$ is half-way between $n = 14$ and $n = 16$) and adding to the latter. Thus

$$r_{\text{rho}(0.05)} = 0.506 + \frac{(0.544 - 0.506)}{2} = 0.525.$$

Since our obtained r_{rho} of 0.65 exceeds the critical value at the 0.05 level, we may conclude that the population value of the Spearman correlation coefficient from which the sample was drawn is greater than 0.00.

CHAPTER SUMMARY

We have seen that if we take a number of samples from a given population, then

1. the distribution of sample means tends to be normal,
2. the mean of these sample means ($\mu_{\overline{X}}$) is equal to the mean of the population (μ), and
3. the standard error of the mean ($\sigma_{\overline{X}}$) is equal to $\sigma/\sqrt{N}$. As N increases, the variability decreases.

We used these relationships in the testing of hypotheses (for example, $\mu = \mu_0$), when the standard deviation of a population was known, employing the familiar z-statistic and the standard normal curve.

When σ is not known, we demonstrated the use of sample statistics to estimate these parameters. We used these estimates of the parameters to test hypotheses, employing the Student t-ratio and the corresponding sampling distributions. We compared these t-distributions, which vary as a function of degrees of freedom (df), with the standard normal curve.

We employed the t-ratio as a basis for establishing confidence intervals.

Finally, we demonstrated the test of significance for the Pearson r and the Spearman r_{rho}, one-sample case.

Terms to Remember:

Standard error of the mean
Central-limit theorem
Law of large numbers
Critical region
Unbiased estimate of a parameter
$\hat{s}^2$ *and* $\hat{s}$
t-distributions
Student's t-ratio
degrees of freedom (df)
Critical values of t
Point estimation
Interval estimation
95% Confidence limits
95% Confidence interval
99% Confidence limits
99% Confidence interval
z_r
Z_r
ρ

EXERCISES

1. Describe what happens to the distribution of sample means when you
   a) increase the size of each sample,
   b) increase the number of samples.
2. Explain why the standard deviation of a sample will usually underestimate the standard deviation of a population. Give an example.
3. Given that $\overline{X} = 24$ and $s = 4$ for $N = 15$, use the t-distribution to find
   a) the 95% confidence limits for μ,
   b) the 99% confidence limits for μ.
4. Given that $\overline{X} = 24$ and $s = 4$ for $N = 121$, use the t-distribution to find
   a) the 95% confidence limits for μ,
   b) the 99% confidence limits for μ.
   Compare the results with Problem 3.
5. An instructor gives his class an examination which, as he knows from years of experience, yields $\mu = 78$ and $\sigma = 7$. His present class of 22 obtains a mean of 82. Is he correct in assuming that this is a superior class? Employ $\alpha = 0.01$, two-tailed test.

6. An instructor gives his class an examination which, as he knows from years of experience, yields $\mu = 78$. His present class of 22 obtains $\overline{X} = 82$ and $s = 7$. Is he correct in assuming that this is a superior class? Employ $\alpha = 0.01$, two-tailed test.

7. Explain the difference between Problems 5 and 6. What test statistic is employed in each case and why? Why is the decision different? Generalize: What is the effect of knowing σ upon the likelihood of a type II error?

8. Superintendent X, of Zody school district, claims that the children in his district are brighter, on the average, than the general population of students. In order to determine the I.Q. of school children in the district, a study was conducted. The results were as follows.

Test scores	Test scores
105	115
109	103
115	110
112	125
124	99

The mean of the general population of school children is 106. Set this up in formal statistical terms (that is, H_0, H_1, etc.), and draw the appropriate conclusions. Employ a one-tailed test, $\alpha = 0.05$.

9. For a particular population with $\mu = 28.5$ and $\sigma = 5.5$, what is the probability that, in a sample of 100, the $\overline{X}$ will be
   a) equal to or less than 30.0?
   b) equal to or less than 28.0?
   c) equal to or more than 29.5?
   d) between 28.0 and 29.0?

10. Given that $\overline{X} = 40$ for $N = 24$ from a population in which $\sigma = 8$, find
    a) the 95% confidence limits for μ,
    b) the 99% confidence limits for μ.

11. It is axiomatic that when pairs of individuals are selected at random and the intelligence test scores of the first members of the pairs are correlated with the second members, $\rho = 0.00$.
    a) Thirty-nine pairs of siblings are randomly selected and an $r = +0.27$ is obtained between members of the pairs for intelligence. Are siblings more alike in intelligence than unrelated individuals?
    b) A study of 28 pairs of monozygotic twins yields $r = +0.91$ on intelligence test scores. What do you conclude?

12. Overton University claims that because of its superior facilities and close faculty supervision, its students complete the Ph.D program earlier than is usual. They base this assertion on the fact that the national mean age for completion is 32.11, whereas the mean age of their 26 Ph.D.'s is 29.61 with $s = 6.00$. Test the validity of their assumption.

13. Employing the data in the above problem, find the interval within which you are 95% confident that the true population mean (average age for Ph. D.'s at Overton University) probably falls.

14. A sociologist asserts that the average length of courtship is longer before a second marriage than before a first. He bases this assertion on the fact that the average for first marriages is 265 days, whereas the average for second marriages (e.g., his 626 subjects) is 268.5 days, with $s = 50$. Test the validity of his assumption.

15. Employing the data in the above problem, find the interval within which you are 99% confident that the true population mean (average courtship days for a second marriage) probably falls.

16. Random samples of size 2 are selected from the following finite population of scores: 1, 3, 5, 7, 9, and 11.
    a) Calculate the mean and standard deviation of the population.
    b) Construct a histogram showing the sampling distribution of means when $n = 2$. Employ sampling *without* replacement.
    c) Construct a histogram showing the means of all possible samples that can be drawn employing sampling *with* replacement.

17. Employing (b) in Problem 16, answer the following: Selecting a sample with $n = 2$, what is the probability that:
    a) A mean as high as 10 will be obtained?
    b) A mean as low as 2 will be obtained?
    c) A mean as deviant as 8 will be obtained?
    d) A mean as low as 1 will be obtained?

18. Employing (c) in Problem 16, answer the following: Selecting a sample with $n = 2$, what is the probability that:
    a) A mean as high as 10 will be obtained?
    b) A mean as low as 2 will be obtained?
    c) A mean as deviant as 8 will be obtained?
    d) A mean as low as 1 will be obtained?

19. A stock analyst claims that he has an unusually accurate method for forecasting price gains of listed common stock. During a given period, stocks he advocated showed the following price gains: \$1.25, \$2.50, \$1.75, \$2.25, \$3.25, \$3.00, \$2.00, \$2.00. During the same period, the market as a whole showed a mean price gain of \$1.83. Set up and test the H_0 that the stocks he selected have been randomly selected from the population of stock gains during the specified period.

20. In a test of a gasoline additive, a group of carefully engineered automobiles were run at a testing site under rigorously supervised conditions. The number of miles obtained in a single gallon of gasoline were: 15, 12, 13, 16, 17, 11, 14, 15, 13, 14. Thousands of prior trials with the same gasoline minus the additive had yielded the expectation of 12.62 miles per gallon. Using $\alpha = 0.05$, can it be concluded that the additive improved gasoline mileage?

21. A restaurant owner ranked his 17 waiters in terms of their speed and efficiency on the job. He correlated these ranks with the total amount of tips each of these

waiters received for a one-week period. The obtained $r_{\text{rho}} = 0.438$. What do you conclude?

22. The owner of a car-leasing company ranked 25 of his customers on their neatness and general care of their rented car during a three-month period. He correlated these ranks with the number of miles each customer drove during this same period. The obtained $r_{\text{rho}} = -0.397$. Employing $\alpha = 0.05$, two-tailed test, what do you conclude?
23. As a requirement for admission to Blue Chip University, a candidate must take a standardized entrance examination. The correlation between performance on this examination and college grades is 0.43.
    a) The director of admissions claims that a better way to predict college success is by using high school averages. To test his claim, he randomly selects 52 students and correlates their college grades with their high school averages. He obtains $r = 0.54$. What do you conclude?
    b) The director's assistant constructs a test which he claims is better for predicting college success than the one currently used. He randomly selects 67 students and correlates their grade point averages with performance on his test. The obtained $r = 0.61$. What do you conclude?
24. What are the statistics used to describe the distribution of a sample? the distribution of a sample statistic?
25. Is s^2 an unbiased estimate of σ^2? Why?
26. Is $\hat{s}^2$ an unbiased estimate of σ^2? Why?
27. What is a confidence interval?
28. Give an example to show the effect of the α-level on the precision of a confidence interval.
29. How do the t-distributions differ from the normal distribution? Are they ever the same?

Statistical Inference With Two Independent Samples 13

13.1 SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN MEANS

What is the effect of drug versus no drug on maze-learning?

Does the recidivism rate of juvenile offenders who are provided with "father figures" differ from those without "father figures?"

Do students in the ungraded classroom differ in performance on standardized achievement tests from students in the straight-age classroom setting?

Each of the above problems involves the comparison of at least two samples. Thus far, for heuristic purposes, we have restricted our examination of hypothesis testing to the one sample case. However, most behavioral research involves the comparison of two or more samples to determine whether or not these samples might have reasonably been drawn from the same population. If the means of two samples differ, must we conclude that these samples were drawn from two different populations?

Recall our previous discussion on the sampling distribution of sample means (Section 12.2). We saw that some variability in the sample statistics is to be expected, even when these samples are drawn from the same population. We were able to describe this variability in terms of the sampling distribution of sample means. To conceptualize this distribution, we imagined drawing an extremely large number of samples, of a fixed N, from a population to obtain the distribution of sample means. In the two sample case, we should imagine drawing pairs of samples, finding the difference between the means of each pair, and obtaining a distribution of these differences. The resulting distribution would be the *sampling distribution of the difference between means.*

To illustrate, imagine that we randomly draw (with replacement) two samples at a time from the population described in Table O in which $\mu = 5.00$ and $\sigma = 0.99$. For illustrative purposes, let us draw two cases for the first sample (that is, $n_1 = 2$), and three cases for the second sample (that is, $n_2 = 3$). For example, we might draw scores of 5, 6 for our first sample and scores of 4, 4, 7 for our second sample. Thus, since $\overline{X}_1 = 5.5$ and $\overline{X}_2 = 5.0$, $\overline{X}_1 - \overline{X}_2 = 0.5$.

Now suppose we continue to draw samples of $n_1 = 2$ and $n_2 = 3$ until we obtain an indefinitely large number of pairs of samples. If we calculate the differences between these pairs of sample means and treat each of these differences as a raw score, we may set up a frequency distribution of these differences.

Intuitively, what might we expect this distribution to look like? Since we are selecting pairs of samples at random from the *same* population, we would expect a normal distribution with a mean of zero.

Going one step further, we may describe the distribution of the difference between pairs of sample means, even when these samples are *not* drawn from the same population. It will be a normal distribution with a mean ($\mu_{\overline{X}_1-\overline{X}_2}$) equal to $\mu_1 - \mu_2$ and a standard deviation ($\sigma_{\overline{X}_1-\overline{X}_2}$ referred to as the *standard error of the difference between means*) equal to $\sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}}$.

Thus, the sampling distribution of the statistic

$$z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{X}_1-\overline{X}_2}}$$

is normal, and therefore permits us to employ the standard normal curve in the testing of hypotheses.

13.2 ESTIMATION OF $\sigma_{\overline{X}_1-\overline{X}_2}$ FROM SAMPLE DATA

The statistic z is employed only when the population standard deviations are known. Since it is rare that these parameters are known, we are once again forced to estimate the standard error in which we are interested, that is, $\sigma_{\overline{X}_1-\overline{X}_2}$.

If we draw a random sample of n_1 observations from a population with unknown variance and a second random sample of n_2 observations from a population with unknown variance, the standard error of the difference between means may be estimated by

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{s^2_{\overline{X}_1} + s^2_{\overline{X}_2}}, \tag{13.1}$$

or

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}. \tag{13.2}$$

However, if $n_1 = n_2 = n$, formula (14.2) simplifies to

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2 + s_2^2}{n - 1}} \tag{13.3}$$

where $n =$ the number in either sample.

The above formulas (13.1) and (13.2) represent biased estimates of the standard error of the difference between means. However, as n becomes larger,

they approach the unbiased estimate. The unbiased estimate, which assumes that the two samples are drawn from a population with the same variance, pools the sums of squares and degrees of freedom of the two samples to obtain a pooled estimate of the standard error of the difference. Hence

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\left(\frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \tag{13.4}^*$$

However, if $n_1 = n_2 = n$, formula (13.4) simplifies to

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sum x_1^2 + \sum x_2^2}{n(n-1)}}. \tag{13.5}^*$$

In order to obtain the sum of squares for our two groups, we apply formula (6.7) to each sample and obtain

$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n_1}, \qquad \text{and} \qquad \sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n_2}.$$

When n's are equal, formulas (13.1), (13.2), and (13.3) are algebraically identical to formulas (13.4) and (13.5) [pooled estimates of $s_{\overline{X}_1-\overline{X}_2}$]. Thus, for equal n's, any of the five formulas may be employed, depending upon computational ease.

13.3 TESTING STATISTICAL HYPOTHESES: STUDENT'S t

The statistic employed in the testing of hypotheses, when population standard deviations are not known, is the familiar t-ratio

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_{\overline{X}_1-\overline{X}_2}}, \tag{13.6}$$

in which $(\mu_1 - \mu_2)$ is the expected value as stated in the null hypothesis.† The Student t-ratio requires an unbiased estimate of $\sigma_{\overline{X}_1-\overline{X}_2}$. Therefore, formula (13.3), (13.4), or (13.5) is used to obtain the value of $s_{\overline{X}_1-\overline{X}_2}$.

---

* This formula can be presented in terms of raw scores. Refer to the endpapers of this book.

† The most common null hypothesis tested is that both samples come from the same population, that is, $(\mu_1 - \mu_2) = 0$.

You will recall that Table C provides the critical values of t required for significance at various levels of α. Since the degrees of freedom for each sample is $n_1 - 1$ and $n_2 - 1$, the total df in the two sample case is $n_1 + n_2 - 2$.

13.3.1 Illustrative Problem: Student's *t*

A researcher wants to determine whether or not a given drug has any effect on the scores of human subjects performing a task of psychomotor coordination. Nine subjects in group 1 (experimental group) receive an oral administration of the drug prior to being tested. Ten subjects in group 2 (control group) receive a placebo at the same time.

Table 13.1

Scores of two groups of subjects on a test of psychomotor coordination

Group 1 Experimental		Group 2 Control	
X_1	X_1^2	X_2	X_2^2
12	144	21	441
14	196	18	324
10	100	14	196
8	64	20	400
16	256	11	121
5	25	19	361
3	9	8	64
9	81	12	144
11	121	13	169
		15	225
$\sum$ 88	996	$\sum$ 151	2445
$n_1 = 9$		$n_2 = 10$	
$\overline{X}_1 = 9.778$		$\overline{X}_2 = 15.100$	

Let us set up this problem in formal statistical terms. The results of the experiment are shown in Table 13.1.

1. *Null hypothesis* (H_0): There is no difference between the population means of the drug group and the no-drug group on the test of psychomotor coordination, that is, $\mu_1 = \mu_2$, or $\mu_1 - \mu_2 = 0$.
2. *Alternative hypothesis* (H_1): There is a difference between the population means of the two groups on the test of psychomotor coordination. Note that our alternative hypothesis is nondirectional. Consequently, a two-tailed test of significance will be employed, that is, $\mu_1 \neq \mu_2$.

3. *Statistical test:* Since we are comparing two sample means presumed to be drawn from normally distributed populations with equal variances, the Student t-ratio two-sample case, is appropriate.
4. *Significance level:* $\alpha = 0.05$.
5. *Sampling distribution:* The sampling distribution is the Student t-distribution with df $= n_1 + n_2 - 2$, or $9 + 10 - 2 = 17$.
6. *Critical region:* $|t| \geq 2.110$. Since H_1 is nondirectional, the critical region consists of all the values of $t \geq 2.110$ and $t \leq -2.110$.

Since $n_1 \neq n_2$ and the population variances are assumed to be equal, we shall employ formula (13.4) to estimate the standard error of the difference between means. The sum of squares for group 1 is

$$\Sigma x_1^2 = 996 - \frac{(88)^2}{9} = 135.56.$$

Similarly, the sum of squares for group 2 is

$$\Sigma x_2^2 = 2445 - \frac{(151)^2}{10} = 164.90.$$

The value of t in the present problem is

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\dfrac{\Sigma x_1^2 + \Sigma x_2^2}{n_1 + n_2 - 2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

$$= \frac{(9.778 - 15.100) - 0}{\sqrt{\left(\dfrac{135.56 + 164.90}{17}\right)\left(\dfrac{1}{9} + \dfrac{1}{10}\right)}} = \frac{-5.322}{1.93} = -2.758.$$

Decision: Since the obtained t falls within the critical region (that is, $|-2.758| > 2.110$, or $-2.758 < -2.110$), we reject H_0. The negative t-ratio simply means that the mean for group 2 is greater than the mean for group 1. When referring to Table C, we ignore the sign of the obtained t-ratio.

13.4 THE t-RATIO AND HOMOGENEITY OF VARIANCE

The assumptions underlying the use of the t-distributions are as follows:

1) The sampling distribution of the difference between means is normally distributed.

2) Estimated $\sigma_{\overline{X}_1-\overline{X}_2}$(that is, $s_{\overline{X}_1-\overline{X}_2}$) is based on the unbiased estimate of the population variance.

3) Both samples are drawn from populations whose variances are equal. This assumption is referred to as *homogeneity of variance.*

Occasionally, for reasons which may not be very clear, the scores of one group may be far more widely distributed than the scores of another group. This may indicate that we are sampling two different distributions, but the critical question becomes: Two different distributions of what? . . . means or variances?

To determine whether or not two variances differ significantly from one another, we must make reference to yet another distribution: the F-distribution. Named after R. A. Fisher, the statistician who first described it, the F-distribution is unlike any other we have encountered in the text; it is tridimensional in nature. To employ the F Table (Table D), we must begin at the entry stating the number of degrees of freedom of the group with the smaller variance, and move down the column until we find the entry for the number of degrees of freedom of the group with the larger variance. At that point, we will find the critical value of F required for rejecting the null hypothesis of no difference in variances. F, itself, is defined as follows:

$$F = \frac{\hat{s}^2 \text{ (larger variance)}}{\hat{s}^2 \text{ (smaller variance)}}. \tag{13.7}$$

In the preceding sample problem, the variance for group 1 is 135.56/8 or 16.94; and for group 2, 164.9/9 or 18.32. The F-ratio becomes

$$F = 18.32/16.94 = 1.08, \qquad \text{df} = 9/8.$$

Referring to Table D, under 9 and 8 df, we find that an F-ratio of 3.39 or larger is required for significance at the 0.05 point or the 0.10 level. If we desire to employ $\alpha = 0.05$, we refer to Table D_1 which provides critical values for F at the 0.025 point or the 0.05 level. Referring to this table, under 9 and 8 df, we find that an $F \geq 4.36$ is significant at the 0.05 level. We may therefore conclude that it is reasonable to assume that both samples were drawn from a population with the same variances.

What if we found a significant difference in variances? Would it have increased our likelihood of rejecting the null hypothesis of no difference between means? Probably not. If anything, a significant difference in variances would have lowered the likelihood of rejecting the null hypothesis. Why, then, do we concern ourselves with an analysis of the variances? Frequently, a significant difference in variances (particularly, when the variance of the experimental group is significantly greater than that of the control group) is indicative of a

dual effect of the experimental conditions. A larger variance indicates more extreme scores *at both ends* of a distribution. The alert researcher will seize upon these facts as a basis for probing into the possibility of dual effects. For example, years ago the experimental question, "Does anxiety improve or hinder performance on complex psychological tasks?," was thoroughly studied with rather ambiguous results.

More recently, we have come to recognize that anxiety-induced conditions have dual effects, depending on a host of factors, e.g., personality variables. An increase in anxiety causes some individuals to become better oriented to the task at hand, while increases in anxiety cause others to "blow up," so to speak. A study of the variances of our experimental groups may facilitate the uncovering of such interesting and theoretically important dual effects.

CHAPTER SUMMARY

We have seen that if we take a number of pairs of samples either from the same population or from two different populations, then

1. the distribution of differences between pairs of sample means tends to be normal,
2. the mean of these differences between means ($\mu_{\overline{X}_1-\overline{X}_2}$) is equal to the difference between the population means, that is, $\mu_1 - \mu_2$,
3. the standard error of the difference between means ($\sigma_{\overline{X}_1-\overline{X}_2}$) is equal to $\sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}}$.

We presented several different formulas for estimating $\sigma_{\overline{X}_1-\overline{X}_2}$ from sample data. Employing estimated $\sigma_{\overline{X}_1-\overline{X}_2}$ (that is, $s_{\overline{X}_1-\overline{X}_2}$), we demonstrated the use of Student's t to test hypotheses in the two-sample case.

An important assumption underlying the use of the t-distributions is that both samples are drawn from populations with equal variances. Although failure to find homogeneity of variance will probably not seriously affect our interpretations, the fact of heterogeneity of variance may have important theoretical implications.

Terms to Remember:

Sampling distribution of the difference between means
Standard error of the difference between means
Student's t-ratio, two-sample case
F-ratio
Homogeneity of variance
Heterogeneity of variance

EXERCISES

1. Two statistics classes of 25 students each obtained the following results on the final examination: $\overline{X}_1 = 82$, $\sum x_1^2 = 384.16$; $\overline{X}_2 = 77$, $\sum x_2^2 = 1536.64$. Test the hypothesis that the two classes are equal in ability, employing $\alpha = 0.01$.

2. In an experiment on the effects of a particular drug on the number of errors in maze-learning behavior of rats, the following results were obtained:

Drug group	Placebo group
$\sum X_1 = 324$	$\sum X_2 = 256$
$\sum X_1^2 = 6516$	$\sum X_2^2 = 4352$
$n_1 = 18$	$n_2 = 16$

Set this experiment up in formal statistical terms, employing $\alpha = 0.05$, and draw the appropriate conclusions concerning the effect of the drug on errors.

3. On a psychomotor task involving two target sizes, the following results were obtained:

Group 1	Group 2
9	6
6	7
8	7
8	9
9	8

Set this experiment up in formal statistical terms, employing $\alpha = 0.05$, and draw the appropriate conclusions.

4. A study was undertaken to determine whether or not the acquisition of a response is influenced by a drug. The criterion variable was the number of trials required to master the task (X_1 is the experimental group and X_2 is the control group).

X_1	6	8	14	9	10	4	7	
X_2	4	5	3	7	4	2	1	3

a) Set this study up in formal statistical terms, and state the appropriate conclusions, employing $\alpha = 0.01$.
b) Is there evidence of heterogeneity of variance?

5. Given two normal populations,

$$\mu_1 = 80, \quad \sigma_1 = 6; \qquad \mu_2 = 77, \quad \sigma_2 = 6.$$

If a sample of 36 cases is drawn from population 1 and a sample of 36 cases from population 2, what is the probability that

a) $\overline{X}_1 - \overline{X}_2 \geq 5$?
b) $\overline{X}_1 - \overline{X}_2 \geq 0$?
c) $\overline{X}_1 - \overline{X}_2 \leq 0$?
d) $\overline{X}_1 - \overline{X}_2 \leq -5$?

6. Assuming the same two populations as in Problem 5, calculate the probability that $\overline{X}_1 - \overline{X}_2 \geq 0$, when

a) $n_1 = n_2 = 4$,
b) $n_1 = n_2 = 9$,
c) $n_1 = n_2 = 16$,
d) $n_1 = n_2 = 25$.

7. Graph the above probabilities as a function of n. Can you formulate any generalization about the probability of finding a difference in the correct direction between sample means (that is, $\overline{X}_1 - \overline{X}_2 \geq 0$, when $\mu_1 > \mu_2$) as a function of n?

8. A gasoline manufacturer runs tests to determine the relative performance of automobiles employing two different additives. The results are as follows (expressed in terms of miles per gallon of gasoline).

   Additive 1: 12, 17, 15, 13, 11, 10, 14, 12
   Additive 2: 16, 14, 18, 19, 17, 13, 11, 18

   Set up and test the appropriate null hypothesis.

9. Each of two market analysts claims that his ability to forecast price gains in common stock is better than his rival. Over a specified period, each selected ten common stocks that he predicted would show a gain. The results were as follows:

   Analyst 1: $1.25, $2.50, $1.75, $2.25, $2.00, $1.75, $2.25, $1.00, $1.75, $2.00
   Analyst 2: $1.25, $.75, $1.00, $1.50, $2.00, $1.75, $.50, $1.50, $.25, $1.25

   Set up and test the appropriate null hypothesis.

10. A toothpaste manufacturer claims that children brushing their teeth daily with his product (brand A) will have fewer cavities than children employing brand X. In a carefully supervised study, the number of cavities in a sample of children using his toothpaste was compared with the number of cavities among children using brand X. The results were as follows:

    Brand A: 1, 2, 0, 3, 0, 2, 1, 4, 2, 3, 1, 2, 1, 1
    Brand X: 3, 1, 2, 4, 1, 5, 2, 0, 5, 6, 3, 2, 4, 3

    Test the manufacturer's claim.

11. The random samples of fifteen manufacturing concerns with total assets under $5,000,000 showed an average after-tax profit of 2.2% of sales and a standard deviation of 0.5%. A random sample of 12 manufacturing concerns with assets between $5,000,000 and $10,000,000 yielded an average after-tax profit of 2.5% of sales and a standard deviation of 0.6%. Is the difference attributable to chance variations which will occur whenever we base our conclusion on samples? Use the 0.05 significance level.

12. A college maintains that it has made vast strides in raising the standards of admission for its entering freshmen. It cites the fact that the mean high school average of 80 entering freshmen was 82.53 for last year with a standard deviation of 2.53. For the present year these statistics are 83.04 and 2.58, respectively, for 84 entering freshmen. Set up and test the appropriate null hypothesis, employing $\alpha = 0.01$.

13. A training director in a large industrial firm claims that employees taking his training course perform better on the job than those not receiving training. Of 30 more recently hired employees, 15 are randomly selected to receive training. The remaining 15 are employed as controls. Six months later, on-the-job test evaluations yield the following statistics for the training group: $\overline{X} = 24.63$, $s = 3.53$. The controls obtain $\overline{X} = 21.45$ and $s = 4.02$. Set up and test the appropriate null hypothesis, employing $\alpha = 0.05$.

14. A publisher claims that students who receive instruction in mathematics based on his newly constructed text will score at least five points higher on end-of-term

grades than those instructed using the previous text. Thirty-six students are randomly assigned to two classes: the experimental group employs the new text for instruction, and the control group uses the previous text. Students in the experimental group achieve $\overline{X} = 83.05$ and $s = 6.04$ as final grades, whereas the controls obtain $\overline{X} = 76.85$ and $s = 5.95$.

Set up and test the appropriate null hypothesis employing $\alpha = 0.01$, one-tailed test. [*Note:* Remember that the numerator in the test statistic is: $(\overline{X}_1 - \overline{X}_2) - (u_1 - u_2)$.]

15. A manufacturer of carpets is considering the replacement of his looms with new machines which are expected to produce carpets at a higher rate per hour. His cost analyst informs him that the exchange is inadvisable unless the new looms produce at a rate that is 10% greater than the old looms. In a test involving 12 old and 10 new looms, the mean number of carpets produced in one hour was found to be 12.30 and 14.45 respectively with standard deviations of 3.45 and 2.96. Employing $\alpha = 0.01$, one-tailed test, decide whether or not the manufacturer should make the change-over to the new looms.

16. If we found a significant difference between means at the 5% level of significance, it would also be true that (true or false):

    a) This difference is significant at the 1% level of significance.
    b) This difference is significant at the 10% level of significance.
    c) The difference observed between means is the true difference.

For Problems 17 through 21, the following two finite populations of scores are given:

Population 1: 2, 4, 6, 8
Population 2: 1, 3, 5, 7

17. Random samples of size 2 are selected (*without replacement*) from each population. Construct a histogram showing the sampling distribution of differences between the sample means.

18. What is the probability that:

    a) $\overline{X}_1 - \overline{X}_2 \geq 0$? b) $\overline{X}_1 - \overline{X}_2 \geq 1$? c) $\overline{X}_1 - \overline{X}_2 \leq 0$?
    d) $\overline{X}_1 - \overline{X}_2 \leq -1$? e) $\overline{X}_1 - \overline{X}_2 \geq 4$? f) $\overline{X}_1 - \overline{X}_2 \leq -4$?
    g) $\overline{X}_1 - \overline{X}_2 \geq 2$ *or* $\overline{X}_1 - \overline{X}_2 \leq -2$?

19. Calculate the mean and standard deviation of each population.

20. If a sample of 25 cases is drawn from population 1 and a sample of 25 cases from population 2, what is the probability that:

    a) $\overline{X}_1 - \overline{X}_2 \geq 0$? b) $\overline{X}_1 - \overline{X}_2 \geq 1$? c) $\overline{X}_1 - \overline{X}_2 \leq 0$?
    d) $\overline{X}_1 - \overline{X}_2 \leq -1$? e) $\overline{X}_1 - \overline{X}_2 \geq 4$?
    f) $\overline{X}_1 - \overline{X}_2 \geq 1$ *or* $\overline{X}_1 - \overline{X}_2 \leq -1$?
    g) $\overline{X}_1 - \overline{X}_2 \geq 2$ *or* $\overline{X}_1 - \overline{X}_2 \leq -2$?

21. Calculate the probability that $\overline{X}_1 - \overline{X}_2 \leq 0$, when

    a) $n_1 = n_2 = 4$ b) $n_1 = n_2 = 9$
    c) $n_1 = n_2 = 16$ d) $n_1 = n_2 = 36$

Statistical Inference With Correlated Samples

14

14.1 INTRODUCTION

One of the fundamental problems confronting the behavioral scientist is the extreme variability of his data. Indeed, it is *because* of this variability that he is so concerned with the field of inferential statistics.

When an experiment is conducted, data comparing two or more groups are obtained, a difference in some measure of central tendency is found, and then we raise the question: Is the difference of such magnitude that it is unlikely to be due to chance factors? As we have seen, a visual inspection of the data is not usually sufficient to answer this question because there is so much overlapping of the experimental groups. The overlapping, in turn, is due to the fact that the experimental subjects themselves manifest widely varying aptitudes and proficiencies relative to the criterion measure. In an experiment, the score of any subject on the criterion variable may be thought to reflect at least three factors: (1) the subject's ability and/or proficiency on the criterion task; (2) the effects of the experimental variable; and (3) random error due to a wide variety of different causes, e.g., minor variations from time to time in experimental procedures or conditions, or momentary fluctuations in such things as attention span, motivation of the experimental subjects, etc. There is little we can do about *random error* except to maintain as close control over experimental conditions as possible. The *effects of the experimental variable* are, of course, what we are interested in assessing. In most studies the *individual differences among subjects* is, by and large, the most significant factor contributing to the scores and the variability of scores on the criterion variable. Anything we can do to take this factor into account or "statistically remove" its effects will improve our ability to estimate the effects of the experimental variable on the criterion scores. This chapter is concerned with a technique that is commonly employed to accomplish this very objective: the employment of correlated samples.

14.2 STANDARD ERROR OF THE DIFFERENCE BETWEEN MEANS FOR CORRELATED GROUPS

In our earlier discussion of Student's t-ratio, we presented the formula for the unpooled estimate of the standard error of the difference between means (13.1) as

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{s^2_{\overline{X}_1} + s^2_{\overline{X}_2}}.$$

Actually, this is not the most general formula for the standard error of the difference. The most general formula is

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{s^2_{\overline{X}_1} + s^2_{\overline{X}_2} - 2rs_{\overline{X}_1}s_{\overline{X}_2}}. \tag{14.1}$$

We drop the last term whenever our sample subjects are assigned to experimental conditions at random for the simple reason that when scores are paired at random, the correlation between the two samples will average zero. Any observed correlation will be spurious since it will represent a chance association. Consequently, when subjects are assigned to experimental conditions at random, the last term reduces to zero (since $r = 0$).

However, there are many experimental situations in which we do not assign our experimental subjects at random. Most of these situations can be placed in one of two classes.

1. **Before-after design.** A reading on the *same* subjects is taken both before and after the introduction of the experimental variable. It is presumed that each individual will remain relatively consistent with himself. Thus there will be a correlation between the before sample and the after sample.

2. **Matched group design.** Individuals in both experimental and control groups are matched on some variable known to be correlated to the criterion or dependent variable. Thus, if we wanted to determine the effect of some drug on learning the solution to a mathematical problem, we might match individuals on the basis of I.Q. estimates, amount of mathematical training, grades in statistics, or performance on other mathematics problems. Such a design has two advantages:

a) It ensures that the experimental groups are "equivalent" in initial ability.
b) It permits us to take advantage of the correlation based on initial ability and allows us, in effect, to remove one source of error from our measurements.

To understand the advantage of employing correlated samples, let us look at a sample problem and calculate the standard error of the difference between means using formula (13.1) based upon unmatched groups and formula (14.1) which takes the correlation into account. Table 14.1 presents data for two groups of subjects matched on a variable known to be correlated with the criterion variable. The members comprising each pair are assigned at random to the experimental conditions.

Table 14.1

Scores of two groups of subjects in an experiment employing matched group design (hypothetical data)

Matched pairs	X_1	X_1^2	X_2	X_2^2	X_1X_2
A	2	4	4	16	8
B	3	9	3	9	9
C	4	16	5	25	20
D	5	25	6	36	30
	$\sum$ 14	54	18	86	67

The following steps are employed in the calculation of the standard error of the difference between means for *unmatched* groups.

Step 1. The sum of squares for group 1 is

$$\sum x_1^2 = 54 - (14)^2/4 = 5.$$

Step 2. The standard deviation for group 1 is

$$s_1 = \sqrt{5/4} = 1.1180.$$

Step 3. The standard error of the mean for group 1 is

$$s_{\overline{X}_1} = s_1/\sqrt{n_1 - 1} = 1.1180/\sqrt{3} = 0.6455.^*$$

Step 4. Similarly, the standard error of the mean for group 2 is

$$s_{\overline{X}_2} = 0.6455.$$

Step 5. The standard error of the difference between means for independent groups is

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{s_{\overline{X}_1}^2 + s_{\overline{X}_2}^2} = 0.91.$$

To calculate the standard error of the difference between means for *matched* groups, the following steps are employed.

* The standard error of the mean may be obtained directly by employing formula (12.8), that is,

$$s_{\overline{X}_1} = \sqrt{\frac{\sum x_1^2}{n_1(n_1 - 1)}} = \sqrt{\frac{5}{4(3)}} = 0.6455.$$

Step 1. Employing formula (8.2),* we find that the correlation between the two groups is

$$r = \frac{\sum x_1 x_2}{\sqrt{(\sum x_1^2)(\sum x_2^2)}} = \frac{4}{\sqrt{(5)(5)}} = 0.80.$$

Step 2. The standard error of the difference between means for matched groups is

$$\begin{aligned} s_{\overline{X}_1 - \overline{X}_2} &= \sqrt{s_{\overline{X}_1}^2 + s_{\overline{X}_2}^2 - 2r s_{\overline{X}_1} s_{\overline{X}_2}} \\ &= \sqrt{0.4167 + 0.4167 - 2(0.80)(0.6455)(0.6455)} = 0.41. \end{aligned}$$

You will note that formula (14.1), which takes correlation into account provides a markedly reduced error term for assessing the significance of the difference between means. In other words, it provides a more sensitive test of this difference and is more likely to lead to the rejection of the null hypothesis when it is false. In the language of inferential statistics, it is a more *powerful* test. We shall discuss the concept of power more fully in Chapter 16. Of course, the greater power, or sensitivity, of formula (14.1) is directly related to our success in matching subjects on a variable that is correlated with the criterion variable. When r is large, $s_{\overline{X}_1 - \overline{X}_2}$ will be correspondingly small. As r approaches zero, the advantage of employing correlated samples becomes progressively smaller.

Balanced against the increased sensitivity of the standard error of the difference between means when r is large, is the loss of degrees of freedom. Whereas the number of degrees of freedom for unmatched samples is $n_1 + n_2 - 2$, the number of degrees of freedom when correlated samples are employed is the number of pairs minus one $(n - 1)$. This difference can be critical when the number of degrees of freedom is small, since, as we saw in Section 12.5.1, larger t-ratios are required for significance when the degrees of freedom are small.

14.3 THE DIRECT-DIFFERENCE METHOD: STUDENT'S *t*-RATIO

Fortunately, it is not necessary to actually determine the correlation between samples in order to find $s_{\overline{X}_1 - \overline{X}_2}$. Another method is available which permits the direct calculation of the standard error of the difference. We shall refer to this method as the *direct-difference method* and represent it symbolically as $s_{\overline{D}}$.

In brief, the direct-difference method consists of finding the differences between the criterion scores obtained by each pair of matched subjects, and treating these differences as if they were raw scores. The null hypothesis is that the obtained mean of the difference scores ($\sum D/N$, symbolized as $\overline{D}$) comes from a population in which the mean difference (μ_D) is some specified value. The t-

* Note that x_2 replaces y in formula (8.2).

ratio employed to test $H_0: \mu_D = 0$ is

$$t = \frac{\overline{D} - \mu_D}{s_{\overline{D}}} = \frac{\overline{D}}{s_{\overline{D}}}. \tag{14.2}$$

The raw score formula for calculating the sum of squares of the difference scores is

$$\sum d^2 = \sum D^2 - (\sum D)^2/n \tag{14.3}$$

where D is the difference between paired scores, and d is the deviation of a difference score (D) from $\overline{D}$. It follows then, that the standard deviation of the difference scores is

$$\hat{s}_D = \sqrt{\frac{\sum d^2}{(n-1)}}. \tag{14.4}$$

Furthermore the standard error of the mean difference may be obtained by dividing formula (14.4) by $\sqrt{N}$. Thus

$$s_{\overline{D}} = \sqrt{\frac{\sum d^2}{n(n-1)}} \tag{14.5}$$

or

$$s_{\overline{D}} = \hat{s}_D/\sqrt{n}. \tag{14.6}$$ *

14.3.1 Sample Problem

The directors of a small private college find it necessary to increase the size of classes. A special film, utilizing the most advanced propaganda techniques, such as film shorts, presents the advantages of larger-size classes. The attitude of a group of ten students is assessed before and after the presentation of this film. It is anticipated that more favorable attitudes (i.e., higher scores) will result from exposure to the film.

Let us set up this problem in formal statistical terms.

1. *Null hypothesis* (H_0): There is no difference in the attitudes of students, before and after viewing the film, that is, $\mu_D = 0$.
2. *Alternative hypothesis* (H_1): The attitudes of the students will be more favorable after viewing the film, that is, $\mu_D < 0$. Note that our alternative hypothesis is directional; consequently, a one-tailed test of significance will be employed.
3. *Statistical test:* Since we are employing a before-after design, the Student t-ratio for correlated samples is appropriate.
4. *Significance level:* $\alpha = 0.01$.

---

* An alternative formula for obtaining $s_{\overline{D}}$ is $s_{\overline{D}} = s_D/\sqrt{n-1}$, where $s_D = \sqrt{\sum d^2/n}$.

Table 14.2
Scores of ten subjects in an experiment employing before-after design (hypothetical data)

Subject	Before	After	Difference	
	X_1	X_2	D	D^2
1	25	28	−3	9
2	23	19	4	16
3	30	34	−4	16
4	7	10	−3	9
5	3	6	−3	9
6	22	26	−4	16
7	12	13	−1	1
8	30	47	−17	289
9	5	16	−11	121
10	14	9	5	25
	171	208	−37	511

5. *Sampling distribution:* The sampling distribution is the Student's t-distribution with df $= n - 1$, or $10 - 1 = 9$.
6. *Critical region:* $t_{0.01} \leq -2.821$. Since H_1 predicts that the scores in the *after* condition will be higher than those in the *before* condition, we expect the difference scores to be negative. Therefore, the critical region consists of all values of $t \leq -2.821$.

Table 14.2 presents the results of this experiment. The following steps are employed in the direct-difference method.

Step 1. The sum of squares of the difference scores is

$$\sum d^2 = 511 - (-37)^2/10 = 374.10.$$

Step 2. The standard error of the mean difference is

$$s_{\overline{D}} = \sqrt{374.10/10(9)} = 2.04.$$

Step 3. The value of $\overline{D}$ is $\overline{D} = -37/10 = -3.70$. (To check the accuracy of $\sum D$ we subtract $\sum X_2$ from $\sum X_1$, that is, $\sum X_1 - \sum X_2 = \sum D$, $171 - 208 = -37$.)

Step 4. The value of t in the present problem is

$$t = \overline{D}/s_{\overline{D}} = -3.70/2.04 = -1.81.$$

Decision: Since the obtained t does not fall within the critical region (that is, $-1.81 > t_{0.01}$), we accept H_0.

14.4 SANDLER'S *A*-STATISTIC

In recent years, a psychologist, Joseph Sandler, has demonstrated an extremely simple procedure for arriving at probability values in all situations involving $H_0: \mu_1 - \mu_2 = 0$ for which the Student t-ratio for correlated samples is appropriate. Indeed, since Sandler's statistic, A, is rigorously derived from Student's t-ratio, the probability values are identical with Student's p-values.

The statistic, A, is defined as follows:

$$A = \frac{\text{the sum of the squares of the differences}}{\text{the square of the sum of the differences}} = \frac{\sum D^2}{(\sum D)^2}. \tag{14.7}$$

By making reference to the table of A (Table E) under $n - 1$ degrees of freedom, we can determine whether our obtained A is *equal* to or *less* than the tabled values at various levels of significance.

Let us illustrate the calculation of A from our previous example. It will be recalled that $\sum D^2 = 511$ and $\sum D = -37$. The value of A becomes

$$A = 511/(-37)^2 = 0.373.$$

Referring to Table E under 9 degrees of freedom, we find that an A *equal to or less than* 0.213 is required for significance at the 0.01 level (one-tailed test). Since 0.373 is greater than the tabled value, we accept the null hypothesis. It will be noted that our conclusion is precisely the same as the one we arrived at by employing the Student t-ratio. This is correct, of course, since the two are mathematically equivalent. Since the calculation of A requires far less time and labor than the determination of t, Sandler's A-distribution can replace Student's t whenever correlated samples are employed.

CHAPTER SUMMARY

The most general formula for the standard error of the difference is

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{s^2_{\overline{X}_1} + s^2_{\overline{X}_2} - 2rs_{\overline{X}_1}s_{\overline{X}_2}}.$$

It is obvious that by matching samples on a variable correlated with the criterion variable, we may reduce the magnitude of the standard error of the difference and thereby provide a more sensitive test of the difference between means. The higher the correlation, of course, the greater the reduction in the standard error of the difference.

We demonstrated the use of the direct-difference method for determining the significance of the difference between the means of correlated samples.

Finally, we demonstrated the use of a mathematically equivalent test, the Sandler A-statistic. Due to its computational ease, the Sandler test will unquestionably replace the Student t-ratio when correlated samples are employed.

Terms to Remember:

Before-after design
Matched group design
Direct-difference method
Standard error of the mean difference
Sandler A-statistic

EXERCISES

1. An investigator employs an experimental procedure wherein each subject (S) performs a task which requires the cooperation of a partner. By prearrangement, the partner plays the role of a complaining rejecting teammate. When the task is completed, S is asked to recall any remarks made by the partner. Prior to the experiment, S's have been tested and classified as either secure or insecure in interpersonal relations and then matched on the basis of intelligence.

   The following statistics were derived from data indicating the number of remarks recalled.

Secure S's	Insecure S's
$\overline{X}_1 = 14.2$	$\overline{X}_2 = 12.9$
$s_1 = 2.0$	$s_2 = 1.5$
$n_1 = 17$	$n_2 = 17$

   $r = 0.55$ between number of remarks recalled and intelligence.

   Set this task up in formal statistical terms, employing $\alpha = 0.01$ (two-tailed test) and state the appropriate conclusions.

2. Had we not employed a matched group design (that is, $r = 0.00$) would our conclusion have been any different? Explain. [*Hint:* In calculating $s_{\overline{X}_1-\overline{X}_2}$, employ formula (13.1), (13.2), or (13.3).]

3. Carmen C., manager of a Little League team in the American League, has said that the American League is more powerful than the National League. Determine the validity of this statement from the home run (HR) figures given below. Employ Student's t-ratio for correlated samples *and* Sandler's A, $\alpha = 0.05$.

Final standing in respective league	American league	Number of HR	National league	Number of HR
1	Minnesota	16	Los Angeles	18
2	Chicago	17	San Francisco	11
3	Baltimore	15	Pittsburgh	14
4	Detroit	12	Cincinnati	10
5	Cleveland	11	Milwaukee	12
6	New York	9	Philadelphia	13
7	California	13	St. Louis	8
8	Washington	16	Chicago	10
9	Boston	18	Houston	9
10	Kansas City	14	New York	15

4. In Problem 3, the teams were paired on the basis of final standings in their respective leagues. The matching technique assumes a correlation between final standing and number of homeruns. Is this assumption valid?

5. Perform a Student t-ratio for independent samples on the data in Problem 3. Why is the obtained t-ratio closer to the rejection region than the answer to Problem 3?

6. Experimenter Larry designs a study in which he matches subjects on a variable which he believes to be correlated with the criterion variable. Assuming H_0 to be false, what is the likelihood of a type II error (compared to the use of Student's t-ratio for uncorrelated samples) in the following situations:
   a) The matching variable is uncorrelated with the criterion variable.
   b) The matching variable is highly correlated with the criterion variable.

7. It has often been stated that women have a higher life expectancy than men. Employ $\alpha = 0.05$ to determine the validity of this statement for
   a) white Americans,
   b) nonwhite Americans,
   c) white males compared to nonwhite females.

Expectation of life in the United States*

	WHITE		NONWHITE (Chiefly Negro)			WHITE		NONWHITE (Chiefly Negro)	
Age	Male	Female	Male	Female	Age	Male	Female	Male	Female
0	67.5	74.4	60.9	66.5	11	58.6	65.2	53.5	58.7
1	68.2	74.8	62.9	68.0	12	57.6	64.2	52.6	57.7
2	67.3	73.9	62.1	67.2	13	56.7	63.2	51.6	56.7
3	66.4	73.0	61.2	66.3	14	55.7	62.2	50.7	55.7
4	65.4	72.0	60.3	65.4	15	54.7	61.3	49.7	54.8
5	64.4	71.0	59.3	64.5	16	53.8	60.3	48.8	53.8
6	63.5	70.1	58.4	63.5	17	52.8	59.3	47.8	52.8
7	62.5	69.1	57.4	62.5	18	51.9	58.3	46.9	51.9
8	61.6	68.1	56.5	61.6	19	51.0	57.4	46.0	50.9
9	60.6	67.1	55.5	60.6	20	50.1	56.4	45.1	50.0
10	59.6	66.2	54.5	59.6					

8. Numerous consumer organizations have criticized the automobile industry for employing odometers which show large variations from one instrument to another and from one manufacturer to another. To test whether or not odometers from

* Source: Reader's Digest 1966 Almanac, p. 492. New York: Reader's Digest Association, with permission.

two competing manufacturers may be considered to have been drawn from a common population, eleven different cars were equipped with two odometers each, one from each manufacturer. All automobiles were driven over a measured course of 100 mi and their odometer readings were tabulated. Apply the appropriate test for the significance of the difference between the odometer readings of each manufacturer, employing $\alpha = 0.01$.

	Manufacturer			Manufacturer	
Automobile	A	B	Automobile	A	B
1	104	102	7	97	99
2	112	106	8	107	102
3	103	107	9	100	98
4	115	110	10	104	101
5	99	93	11	108	102
6	104	101			

9. Another complaint by consumer organizations is that the odometers are purposely constructed to overestimate the distance traveled in order to inflate the motorist's estimates of gasoline mileage. As a review of Chapter 12, conduct a one-sample test of $H_0:\mu_0 = 100$ miles for the product of each manufacturer.

10. Runyon (1968) showed that a simple extension of the Sandler A-test may be employed as an algebraically equivalent substitute for the Student t-ratio, one-sample case. The technique consists of subtracting the value of the mean hypothesized under H_0 from each score, summing the differences and squaring $[(\Sigma D)^2]$, squaring the differences and summing $[\Sigma D^2]$, and substituting these values in the formula for the A-statistic, that is, $A = \Sigma D^2/(\Sigma D)^2$. Degrees of freedom equal $n - 1$. Conduct a one-sample test of $H_0:\mu_0 = 100$ miles for the odometers of each manufacturer. Compare the results of this analysis with the results of the foregoing analysis.

11. In a study aimed at determining the effectiveness of a new diet, an insurance company selects a sample of 12 overweight men between age 40 and 50 and obtains their weight measurements both before initiating the diet and sixty days later. Set up and test the appropriate null hypothesis, employing $\alpha = 0.05$.

Subject	Weight		Subject	Weight	
	Before	After		Before	After
1	202	180	7	209	205
2	237	221	8	191	196
3	173	175	9	200	185
4	161	158	10	189	187
5	185	180	11	177	172
6	210	197	12	184	186

12. A difficulty with interpreting the results of Problem 11 is that a control group is lacking. It is possible that a random selection of overweight men who are *not* on a diet will reveal weight losses over a two-month period. A control group, matched in weight with the experimental subjects in Problem 11, demonstrated the following before-after changes.

Subject	Weight		Subject	Weight	
	Before	After		Before	After
1	203	207	7	209	215
2	235	231	8	192	215
3	175	172	9	201	184
4	159	164	10	187	196
5	183	187	11	178	184
6	210	204	12	185	173

Set up and test the appropriate null hypothesis for the control subjects. Employ $\alpha = 0.05$.

13. Employing the "after" weight only for the matched subjects in Problems 11 and 12, test for the significance of the difference between the two conditions, employing $\alpha = 0.05$.

14. Obtain the before-after difference score for each subject in Problems 11 and 12. Conduct a "matched-pairs" analysis of the difference scores, employing $\alpha = 0.05$. Compare the results of this analysis with the foregoing analysis.

15. A large discount house advertises that its prices are lower than its largest competitor. To test the validity of this claim, the prices of 15 randomly selected items are compared. The results are as follows:

Discount house	Competitor	Discount house	Competitor
\$3.77	\$3.95	\$2.99	\$2.95
7.50	7.75	1.98	2.49
4.95	4.99	0.49	0.52
3.18	3.25	5.50	5.62
5.77	5.98	0.99	0.98
2.49	2.39	6.49	6.66
8.77	9.49	5.49	5.55
6.99	6.49		

What do you conclude?

An Introduction to the Analysis of Variance

15

15.1 MULTIGROUP COMPARISONS

We have reviewed the classic design of experiments on several different occasions in this text. The classic study consists of two groups, an experimental and a control. The purpose of statistical inference is to test specific hypotheses, e.g., whether or not both groups could have reasonably been drawn from the same population (see Chapter 13).

Although this classical research design is still employed in many studies, its limitations should be apparent to you. To restrict our observations to two groups, on all occasions, is to overlook the wonderful complexity of the phenomena which the scientist investigates. Rarely do events in nature conveniently order themselves into two groups, an experimental and a control. More commonly, the questions we pose to nature are: Which of several alternative schedules of reinforcement leads to the greatest resistance to experimental extinction? Which of five different methods of teaching the concept of fractions to the primary grades leads to the greatest learning gains? Which form of psychotherapy leads to the greatest incidence of patient recovery?

Obviously, the research design, necessary to provide experimental answers to the above questions, would require comparison of more than two groups. You may wonder: But why should multigroup comparisons provide any obstacles? Can we not simply compare the mean of each group with the mean of every other group and obtain a Student t-ratio for each comparison? For example, if we had four experimental groups, A, B, C, D, could we not calculate Student t-ratios comparing A with B, C and D; B with C and D; and C with D?

If you will think for a moment of the errors in inference, which we have so frequently discussed, you will recall that our greatest concern has been to avoid type I errors. When we establish the region of rejection at the 0.05 level, we are, in effect, acknowledging our willingness to take the risk of being wrong as often as 5% of the time in our rejection of the null hypothesis. Now, what happens when we have numerous comparisons to make? For an extreme example, let us imagine that we have conducted a study involving the calculation of 1000

separate Student t-ratios. Would we be terribly impressed if, say, 50 of the t's proved to be significant at the 0.05 level? Of course not. Indeed, we would probably murmur something to the effect that, "With 1000 comparisons, we would be surprised if we didn't obtain approximately 50 comparisons that are significant *by chance* (i.e., due to predictable sampling error)."

The *analysis of variance* is a technique of statistical analysis which permits us to overcome the ambiguity involved in assessing significant differences when more than one comparison is made. It allows us to answer the question: Is there an overall indication that the experimental treatments are producing differences among the means of the various groups? Although the analysis of variance may be used in the two-sample case (in which event it yields precisely the same probability values as the Student t-ratio), it is most commonly employed when three or more groups are involved. Indeed, it has its greatest usefulness when two or more independent variables are studied. However, in this text we shall restrict our introductory treatment of analysis of variance to several levels of a single independent variable. For purposes of exposing the fundamental characteristics of the analysis of variance, our initial illustrative material will involve the two-sample case.

15.2 THE CONCEPT OF SUMS OF SQUARES

You will recall that we previously defined the unbiased variance estimate as

$$\hat{s}^2 = \frac{\sum x^2}{N - 1}.$$

It will be recalled that, when the deviation of scores from the mean, $(X - \overline{X})$ or x, is large, the variance, and therefore the variability of scores, is also large. When the deviations are small, the variance is correspondingly small.

Now, if we think back to the Student t-ratio for a moment, we shall note that both the numerator and denominator give us some estimate of variability:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}.$$

It will be recalled that the denominator, which we referred to as the standard error of the difference between means, is based on the pooled estimate of the variability within each experimental group, that is,

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

However, the numerator is also a measure of variability, i.e., the variability between means. When the difference between means is large relative to $s_{\overline{X}-\overline{X}_2}$,

the Student t-ratio is large. When the difference between means is small relative to $s_{\overline{X}_1-\overline{X}_2}$, the Student t-ratio is also small.

The analysis of variance consists of obtaining two independent estimates of variance, one based upon variability between groups (*between-group variance*) and the other based upon the variability within groups (*within-group variance*). The significance of the difference between these two variance estimates is provided by Fisher's F-distributions. (We are already familiar with F-distributions from our prior discussion of homogeneity of variance, Section 13.4.) If the *between-group variance* is large (i.e., the difference between means is large) relative to the *within-group variance*, the F-ratio is large. Conversely, if the *between-group variance* is small relative to the *within-group variance*, the F-ratio will be small.

A basic concept in the analysis of variance is the sum of squares. We have already encountered the sum of squares in calculating the standard deviation, the variance, and the standard error of the difference between means. It is simply the numerator in the formula for variance, that is, $\sum x^2$. As you will recall, the raw score formula for calculating the sum of squares is

$$\sum x^2 = \sum X^2 - (\sum X)^2/N.$$

The advantage of the analysis of variance technique is that we can partial the total sum squares ($\sum x^2_{\text{tot}}$) into two components, the *within-group sum squares* ($\sum x^2_W$) and the *between-group sum squares* ($\sum x^2_B$). Before proceeding any further, let us clarify each of these concepts with a simple example.

Imagine that we have completed a study comparing two experimental treatments and obtained the scores listed in Table 15.1.

Table 15.1
Scores of two groups of subjects in a hypothetical experiment

Group 1 X_1	Group 1 X_1^2	Group 2 X_2	Group 2 X_2^2
1	1	6	36
2	4	7	49
5	25	9	81
8	64	10	100
$\sum$ 16	94	32	266

$n_1 = 4, \quad \overline{X}_1 = 4 \quad n_2 = 4, \quad \overline{X}_2 = 8$
$\sum X_{\text{tot}} = 48, \quad N = 8, \quad \overline{X}_{\text{tot}} = 6.$

The mean for group 1 is 4; the mean for group 2 is 8. The overall mean, $\overline{X}_{\text{tot}}$, is $\frac{48}{8}$ or 6. Now, if we were to subtract the overall mean from each score and square, we would obtain the total sum squares:

$$\sum x_{\text{tot}}^2 = \sum(X - \overline{X}_{\text{tot}})^2. \tag{15.1}$$

The alternative raw score formula is

$$\sum x_{\text{tot}}^2 = \sum X_{\text{tot}}^2 - (\sum X_{\text{tot}})^2/N. \tag{15.2}$$

For the data in Table 15.1, the total sum squares is

$$\begin{aligned} \sum x_{\text{tot}}^2 &= 360 - (48)^2/8 \\ &= 360 - 288 = 72. \end{aligned}$$

The within-group sum squares is merely the sum of the sum squares obtained within each group, that is,

$$\sum x_W^2 = \sum x_1^2 + \sum x_2^2 \tag{15.3}$$

$$\begin{aligned} \sum x_1^2 &= \sum X_1^2 - (\sum X_1)^2/n_1 \\ &= 94 - (16)^2/4 = 94 - 64 = 30, \\ \sum x_2^2 &= \sum X_2^2 - (\sum X_2)^2/n_2 \\ &= 266 - (32)^2/4 = 266 - 256 = 10, \\ \sum x_W^2 &= 30 + 10 = 40. \end{aligned}$$

Finally, the between-group sum squares ($\sum x_B^2$) may be obtained by subtracting the overall mean from each group mean, squaring the result, multiplying by the n in each group, and summing across all the groups. Thus

$$\sum x_B^2 = \sum n_i(\overline{X}_i - \overline{X}_{\text{tot}})^2, \tag{15.4}$$

where n_i is the number in the ith group, and $\overline{X}_i$ is the mean of the ith group,

$$\sum x_B^2 = 4(4 - 6)^2 + 4(8 - 6)^2 = 32.$$

The raw score formula for calculating the between-group sum squares is

$$\sum x_B^2 = \sum \frac{(\sum X_i)^2}{n_i} - \frac{(\sum X_{\text{tot}})^2}{N}, \tag{15.5}$$

and

$$\begin{aligned} \sum x_B^2 &= (16)^2/4 + (32)^2/4 - (48)^2/8 \\ &= 64 + 256 - 288 \\ &= 320 - 288 = 32. \end{aligned}$$

It will be noted that the total sum squares is equal to the sum of the between-group sum squares and the within-group sum squares. In other words,

$$\sum x_{\text{tot}}^2 = \sum x_W^2 + \sum x_B^2. \tag{15.6}$$

In the above example, $\sum x_{\text{tot}}^2 = 72$, $\sum x_W^2 = 40$, and $\sum x_B^2 = 32$. Thus $72 = 40 + 32$.

15.3 OBTAINING VARIANCE ESTIMATES

Now, to arrive at variance estimates from between- and within-group sum squares, all we need to do is divide each by the appropriate number of degrees of freedom. The degrees of freedom of the between-group is simply the number of groups (k) minus 1.

$$\mathrm{df}_B = k - 1. \tag{15.7}$$

With two groups ($k = 2$), df $= 2 - 1 = 1$. Thus, our between-group variance estimate for the problem at hand is

$$\hat{s}_B^2 = \sum x_B^2/\mathrm{df}_B = \tfrac{32}{1} = 32, \qquad \mathrm{df} = 1. \tag{15.8}$$

The number of degrees of freedom of the within-groups is the total N minus the number of groups. Thus

$$\mathrm{df}_W = N - k. \tag{15.9}$$

In the present problem, $\mathrm{df}_W = 8 - 2 = 6$ and our within-group variance estimate becomes

$$\hat{s}_W^2 = \sum x_W^2/\mathrm{df}_W = \tfrac{40}{6} = 6.67, \qquad \mathrm{df} = 6. \tag{15.10}$$

Now, all that is left is to calculate the F-ratio and determine whether or not our two variance estimates could have reasonably been drawn from the same population. If not, we shall conclude that the significantly larger between-group variance is due to the operation of the experimental conditions. In other words, we shall conclude that the experimental treatments produced a significant difference in means. The F-ratio, in analysis of variance, is the between-group variance estimate divided by the within-group variance estimate. Symbolically,

$$F = \hat{s}_B^2/\hat{s}_W^2. \tag{15.11}$$

For the above problem our F-ratio is

$$F = 32/6.67 = 4.80, \qquad \mathrm{df} = 1/6.$$

Looking up the F-ratio under 1 and 6 degrees of freedom, Table D, we find that an F-ratio of 5.99 or larger is required for significance at the 0.05 level. For the present problem, then, we cannot reject the null hypothesis.

15.4 FUNDAMENTAL CONCEPTS OF ANALYSIS OF VARIANCE

In these few pages, we have examined all the basic concepts necessary to understand simple analysis of variance. Before proceeding with an example involving three groups, let us briefly review these fundamental concepts.

1. We have seen that in an experiment involving two or more groups, it is possible to identify two different bases for estimating the population variance: the between-group and the within-group.

   a) The between-group variance estimate reflects the magnitude of the difference between and/or among the group means. The larger the difference between means, the larger the between-group variance.

   b) The within-group variance estimate reflects the dispersion of scores within each treatment group. The within-group variance is analogous to $s_{\overline{X}_1-\overline{X}_2}$ in the Student t-ratio. It is often referred to as the error term.

2. The null hypothesis is that the two independent variance estimates may be regarded as estimates of the same population value. In other words, H_0 is that the samples were drawn from the same population, or that $\mu_1 = \mu_2 = \cdots = \mu_k$.

3. The F-ratio consists of the between-group variance estimate divided by the within-group variance estimate. By consulting Table D of the distribution of F we can determine whether or not the null hypothesis of equal population variance can reasonably be entertained. In the event of a significant F-ratio, we may conclude that the group means are not all estimates of a common population mean.

4. In the two-sample case, the F-ratio yields probability values identical to those of the Student t-ratio. Indeed, in the one-degree-of-freedom situation (that is, $k = 2$), $t = \sqrt{F}$ or $t^2 = F$. You may check this statement by calculating the Student t-ratio for the sample problem we have just completed.

15.5 AN EXAMPLE INVOLVING THREE GROUPS

Let us imagine that you have just completed a study concerned with the determination of the effectiveness of three different methods of instruction for the basic principles of arithmetic. Twnety-seven primary grade children were randomly assigned to three equal groups employing one of the methods for achieving randomness described in Section 10.2. Following the completion of their instruction, all children were tested on an "Inventory of Basic Arithmetic." The results of this hypothetical study are presented in Table 15.2.

Table 15.2

Scores of three groups of subjects in a hypothetical experiment

Group 1		Group 2		Group 3	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
4	16	12	144	1	1
5	25	8	64	3	9
4	16	10	100	4	16
3	9	5	25	6	36
6	36	7	49	8	64
10	100	9	81	5	25
1	1	14	196	3	9
8	64	9	81	2	4
5	25	4	16	2	4
$\sum$ 46	292	78	756	34	168

$n_1 = 9, \quad \overline{X}_1 = 5.11 \quad n_2 = 9, \quad \overline{X}_2 = 8.67 \quad n_3 = 9, \quad \overline{X}_3 = 3.78$

$$\sum X_{\text{tot}} = 46 + 78 + 34 = 158,$$
$$\sum X_{\text{tot}}^2 = 292 + 756 + 168 = 1216,$$
$$N = 27.$$

The following steps are employed in a three-group analysis of variance:

Step 1. Employing formula (15.2), the total sum squares is

$$\sum x_{\text{tot}}^2 = 1216 - (158)^2/27 = 291.41.$$

Step 2. Employing formula (15.5) for three groups, the between-group sum squares is

$$\sum x_B^2 = (46)^2/9 + (78)^2/9 + (34)^2/9 - (158)^2/27 = 114.96.$$

Step 3. The within-group sum squares may be obtained by employing formula (15.3) for three groups:

$$\sum x_W^2 = (292 - (46)^2/9) + (756 - (78)^2/9) + (168 - (34)^2/9) = 176.45.$$

You may obtain the within-group sum squares by subtraction, that is,

$$\sum x_W^2 = \sum x_{\text{tot}}^2 - \sum x_B^2 = 291.41 - 114.96 = 176.45.$$

Step 4. The between-group variance estimate is

$$\text{df}_B = k - 1 = 2, \qquad \hat{s}_B^2 = 114.96/2 = 57.48.$$

Step 5. The within-group variance estimate is

$$\mathrm{df}_W = N - k = 24, \qquad \hat{s}_W^2 = 176.45/24 = 7.35.$$

Step 6. Employing formula (15.11) we find that the value of F is

$$F = 57.48/7.35 = 7.82, \qquad \mathrm{df} = 2/24.$$

To summarize these steps, we employ the format shown in Table 15.3.

Table 15.3

Summary table for representing the relevant statistics in analysis of variance problems

Source of variation	Sum squares	Degrees of freedom	Variance estimate*	F
Between-groups	114.96	2	57.48	7.82
Within-groups	176.45	24	7.35	
Total	291.41	26		

* In many texts, the term "mean square" appears in this box. However, the authors prefer the term "variance estimate" since this term accurately describes the nature of the entries in the column.

By employing the format recommended in Table 15.3, you have a final check upon your calculation of sum squares and your assignment of degrees of freedom. Thus, $\sum x_B^2 + \sum x_W^2$ must equal $\sum x_{\text{tot}}^2$. The degrees of freedom of the total are found by

$$\mathrm{df}_{\text{tot}} = N - 1. \tag{15.12}$$

In the present example, the number of degrees of freedom for the total is

$$\mathrm{df}_{\text{tot}} = 27 - 1 = 26.$$

15.6 THE INTERPRETATION OF F

When we look up the F required for significance with 2 and 24 degrees of freedom, we find that an F of 3.40 or larger is significant at the 0.05 level.

Since our F of 7.86 exceeds this value, we may conclude that the three group means are not all estimates of a common population mean. Now, do we stop at this point? After all, are we not interested in determining whether or not one of the three methods of instructing the fundamentals of arithmetic is superior to the other two? The answer to the first question is negative and the answer to the second is affirmative.

The truth of the matter is that our finding an overall significant F-ratio now permits us to investigate specific hypotheses. In the absence of a significant F-ratio, any significant differences between specific comparisons would have to be regarded as suspicious—very possibly representing a chance difference.

Over the past number of years, a large number of tests have been developed which permit the researcher to investigate specific hypotheses concerning population parameters. Two broad classes of such tests exist:

1. *A priori or planned comparisons:* When comparisons are planned in advance of the investigation, an *a priori* test is appropriate. For *a priori* tests, it is not necessary that the overall F-ratio be significant.
2. *A posteriori* comparisons: When the comparisons are not planned in advance, an *a posteriori* test is appropriate.

In the present example, we shall illustrate the use of an *a posteriori* test for making pairwise comparisons among means.

Tukey (1953) has developed such a test which he named the HSD (honestly significant difference) test. To employ this test, the overall F-ratio must be significant.

A difference between two means is significant, at a given α-level, if it equals or exceeds HSD, which is:

$$\text{HSD} = q_\alpha \sqrt{\frac{\hat{s}_W^2}{n}} \tag{15.13}$$

in which $\hat{s}_W^2$ = the within-group variance estimate

n = number of subjects in each condition

q_α = tabled value for a given α-level found in Table P for df_W and k = number of means.

15.6.1 A Worked Example

Let us employ the data from Section 15.5 to illustrate the application of the HSD test. We shall employ $\alpha = 0.05$ for testing the significance of the difference between each pair of means.

Table 15.4
Differences among means

	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$
$\overline{X}_1 = 5.11$	...	3.56	1.33
$\overline{X}_2 = 8.67$	...	...	4.89
$\overline{X}_3 = 3.78$	...	...	...

Step 1. Prepare a matrix showing the mean of each condition and the differences between pairs of means. This is shown in Table 15.4.

Step 2. Referring to Table P under error df $= 24$, $k = 3$ at $\alpha = 0.05$, we find $q_{0.05} = 3.44$.

Step 3. Find HSD by multiplying $q_{0.05}$ by $\sqrt{\dfrac{\hat{s}_W^2}{n}}$. The quantity $\hat{s}_W^2$ is found in Table 15.3 under within-group variance estimate. The n per condition is 9. Thus:

$$\begin{aligned} \text{HSD} &= 3.44\sqrt{\frac{7.35}{9}} \\ &= 3.44\ (0.90) \\ &= 3.10 \end{aligned}$$

Step 4. Referring back to Table 15.4, we find that the differences between $\overline{X}_1$ vs. $\overline{X}_3$ and $\overline{X}_2$ vs. $\overline{X}_3$ both exceed HSD $= 3.10$. We may therefore conclude that these differences are statistically significant at $\alpha = 0.05$.

15.7 WITHIN-GROUP VARIANCE AND HOMOGENEITY

When discussing the Student t-ratio in Chapter 13, we noted that a fundamental assumption underlying the use of the Student t-ratio is that the variances for all groups must be homogeneous, i.e., drawn from the same population of variances. The same assumption holds true for the analysis of variance. In other words, a basic assumption underlying the analysis of variance is that the treatment variances (which, when summed together, makes up $\hat{s}_W^2$) are homogeneous. As with the Student t-ratio, there is a test for determining whether or not the hypothesis of identical variances is tenable. However, it is beyond the scope of this introductory text to delve into Bartlett's test of homogeneity of variances. Application of this test is described in Edwards (1968).

CHAPTER SUMMARY

We began this chapter with the observation that the scientist is frequently interested in conducting studies which are more extensive than the classical two-group design. However, when more than two groups are involved in a study, we increase the risk of making a type I error if we accept, as significant, *any* comparison which falls within the rejection region. In multigroup studies, it is desirable to know whether or not there is an indication of an overall effect of the experimental treatments before we investigate specific hypotheses. The analysis of variance technique provides such a test.

In this chapter we presented a mere introduction to the complexities of analysis of variance. We showed that total sum squares can be partitioned into two component sum squares: the within-group and the between-group. These two component sum squares provide us, in turn, with independent estimates of the population variance. A between-group variance estimate which is large, relative to the within-group variance, suggests that the experimental treatments are responsible for the large differences among the group means. The significance of the difference in variance estimates is obtained by reference to the F-table (Table D).

When the overall F-ratio is found to be statistically significant, we are free to investigate specific hypotheses, employing a multiple-comparison test.

Finally, we pointed out a basic assumption of the analysis of variance technique, i.e., homogeneity of variances among treatment groups.

Terms to Remember:

Analysis of variance
Sum of squares
Variance
Between-group variance
Within-group variance
Total sum of squares
Within-group sum of squares
Between-group sum of squares
Between-group variance estimate
Within-group variance estimate
F-ratio
Homogeneity of variance
A priori or planned comparisons
A posteriori comparisons

EXERCISES

1. Using the following data, derived from the 10-year period 1955–1964, determine whether there is a significant difference, at the 0.01 level, in death rate among the various seasons. (*Note:* Assume death rates for any given year to be independent.)

Winter	Spring	Summer	Fall
9.8	9.0	8.8	9.4
9.9	9.3	8.7	9.4
9.8	9.3	8.8	10.3
10.6	9.2	8.6	9.8
9.9	9.4	8.7	9.4
10.7	9.1	8.3	9.6
9.7	9.2	8.8	9.5
10.2	8.9	8.8	9.6
10.9	9.3	8.7	9.5
10.0	9.3	8.9	9.4

2. Conduct an HSD test, comparing the death rates of each season with every other season. Employ the 0.01 level, two-tailed test, for each comparison.

3. Conduct an analysis of variance on the data in Chapter 13, Problem 3. Verify that, in the two-group condition, $F = t^2$.

4. Manufacturer Pass negotiates contracts with 10 different independent research organizations to compare the effectiveness of his product with that of his leading competitor. A significant difference (0.05 level) in favor of Mr. Pass' product is found in one of the 10 studies. He subsequently advertised that independent research has demonstrated the superiority of his product over the leading competitor. Criticize this conclusion.

5. Various drug companies make the claim that they manufacture an analgesic which releases its active ingredient "faster." A random selection of the products of each manufacturer revealed the following times, in seconds, required for the release of 50% of the analgesic agent. Test the null hypothesis that all the analgesics are drawn from a common population of means.

Brand A	Brand B	Brand C	Brand D
28	34	29	22
19	23	24	31
30	20	33	18
25	16	21	24

6. A consumer organization randomly selects several gas clothes dryers of three leading manufacturers for study. The time required for each machine to dry a standard load of clothes was tabulated. Set up and test the appropriate null hypothesis. Conduct an HSD test comparing each brand with every other brand, employing $\alpha = 0.05$.

Brand A	Brand B	Brand C
42	52	38
36	48	44
47	43	33
43	49	35
38	51	32

7. Automobile tires, selected at random from six different brands, required the following braking distances, in feet, when moving at 25 mi/hr. Set up and test the appropriate null hypothesis. Conduct an HSD test comparing each brand with every other brand, employing $\alpha = 0.01$.

Brand A	Brand B	Brand C	Brand D	Brand E	Brand F
22	25	17	21	27	20
20	23	19	24	29	14
24	26	15	25	24	17
18	22	18	23	25	15

8. If the F-ratio is less than 1.00, what do we conclude?

REVIEW

A. PARAMETRIC TESTS OF SIGNIFICANCE

In Section II.A (Chapters 10–15) you learned the elements of probability theory and their application to the problem of drawing inferences from samples presumed to be selected from normally distributed populations. The problems presented below will allow you to integrate the material you have learned in this section.

1. Assume that the following is a population of scores: 2, 3, 3, 4, 4, 4, 5, 5, 6.

a) Determine the mean and standard deviation of the population.

b) Construct a frequency distribution of means for $N = 2$, employing sampling *with replacement.*

c) Construct a probability distribution of means for $N = 2$.

d) Determine $\sigma_{\overline{X}}$ by use of the following formula:

$$\sigma_{\overline{X}} = \sigma/\sqrt{N}.$$

e) Determine $\sigma_{\overline{X}}$ by direct calculation from the sampling distribution of means.

f) Determine the probability of randomly selecting samples ($N = 2$) with
  i) a sample mean of 2.0,
  ii) a sample mean of at least 5.0,
  iii) a sample mean as rare as 3.0.

2. Assume a second population of scores: 4, 5, 5, 6, 6, 6, 7, 7, 8.

Individual A selects samples (with replacement) from population 1 and individual B selects samples from population 2 (with replacement). They test the null hypothesis that the samples were drawn from a common population, employing $\alpha = 0.05$, two-tailed test. In which of the following examples will they make a type II error? Use only the statistics calculated from the samples.

Do not use the population values.

a)

A	2	3	4	5	6
B	4	5	6	6	8

b)

A	2	3	3	4	4	4	4	4	4	6
B	4	4	4	5	5	5	5	5	6	7

3. Using the data above, Problem 2(a) and (b), assume matching of the scores and test for the significance of differences, employing $\alpha = 0.05$, two-tailed test.

4. Each of three universities claims that it has the brightest students. Each sends ten of its best students to compete in a national contest. The results were as follows:

1	2	3	1	2	3
99	94	74	95	93	99
93	98	99	97	96	71
84	97	98	91	96	70
89	92	99	88	91	79
72	92	97	89	90	83

Set up and test the appropriate null hypothesis.

B. NONPARAMETRIC TESTS OF SIGNIFICANCE

So far we have concerned ourselves with the various facets of the normal probability curve as applied to descriptive and inferential statistics. Statistical tests of inference which make use of the normal probability model are referred to as *parametric tests of significance.*

Many data are collected which either do not lend themselves to analysis in terms of the normal probability curve, or fail to meet the basic assumptions for its application. Consider a study in which the data collected consist of ranks (e.g., ranking students in terms of cooperativeness). The resulting ordinal values are nonquantitative and are, of necessity, distributed in a rectangular fashion. Clearly, the normal probability statistics do not apply. Recent years have seen the development of a remarkable variety of nonparametric tests which may be employed with such data.

> *A nonparametric test of significance is defined as one which makes no assumptions concerning the shape of the parent distribution or population, and accordingly is commonly referred to as a distribution-free test of significance.*

In this final section of the text we shall describe several of the more important nonparametric statistical tests.

Power and Power Efficiency of a Statistical Test

16

16.1 THE CONCEPT OF POWER

In the entire first section of the book, only fleeting references were made to the power and the power efficiency of a statistical test (although they were not identified as such). Before proceeding into nonparametric tests of significance, it is desirable to examine these concepts in more detail.

While discussing type I and type II errors in Section 11.6, we pointed out that the basic conservatism of a scientist causes him to set up a rejection level sufficiently low to make type I errors less frequently than type II errors. In other words, the scientist would rather make the mistake of accepting a false null hypothesis than the mistake of rejecting a true one. However, this conservatism should not be construed to mean that the scientist is happy about the prospect of making type II errors. To the contrary, it is quite likely that many promising research projects have been abandoned because of the failure of the experimenter to reject the null hypothesis when it was actually false.

Now, up to this point in the book, our concern has been to establish a level of significance which will reduce the likelihood of falsely rejecting the null hypothesis. In other words, we have been primarily concerned with avoiding type I rather than type II errors. However, it should be recognized that the ideal statistical test is one which effects some sort of balance between these two types of error. Ideally, we should specify in advance of our study the probability of making both a type I and a type II error. In practice, however, most researchers content themselves with stating only the p-value which they will employ to reject the null hypothesis. As we have seen, this p-value represents the probability of a type I error (that is, α).

When we begin to concern ourselves with effecting a balance between type I and type II errors, we are dealing with the concept of the *power of a test*. The power of a test is defined, simply, as the probability of rejecting the null hypothesis when it is, in fact, false. Symbolically, power is defined as follows:

$$\text{Power} = 1 - \text{probability of a type II error.}$$

If we let β represent the probability of a type II error, the definition of power becomes:

$$\text{Power} = 1 - \beta. \tag{16.1}$$

We can calculate the power of a test only when H_0 is false and we are given the true value of the population mean under H_1.

The following section illustrates the calculation of power.

16.2 CALCULATION OF POWER: ONE-SAMPLE CASE

A psychologist working for a large industrial firm has constructed two aptitude scales which he administers interchangeably to incoming groups of trainees. He knows that the average performance on scale A is 70, and on scale B is 72. Both scales have a standard deviation of 5. He is chagrined to discover that his assistant failed to record which scale was administered to a group of 16 trainees. Scanning the data and noticing a number of low scores, he believes that this sample came from a population in which $\mu = 70$ (i.e., scale A was administered to this group), or H_0: $\mu = \mu_0 = 70$.

As a matter of fact, however, scale B was administered. Thus, since we know that H_0 is false and we know the true value of μ under H_1 ($\mu = \mu_1 = 72$), we may calculate the power of the test (i.e., the probability that he will correctly reject the false null hypothesis).

Let us set up this problem in formal statistical terms.

1. *Null hypothesis* (H_0)*:* The mean of the population from which this sample was drawn equals 70, that is, $\mu = \mu_0 = 70$.
2. *Alternative hypothesis* (H_1)*:* The mean of the population from which this sample was drawn equals 72, that is, $\mu = \mu_1 = 72$.
3. *Statistical test:* Since σ is known, $z = (\overline{X} - \mu_0)/\sigma_{\overline{X}}$ is the appropriate test statistic.
4. *Significance level:* $\alpha = 0.01$ (one-tailed test).
5. *Sampling distribution:* The sampling distribution of the mean is known to be a normal distribution.
6. *Critical region:* $q_{0.01} \geq +2.33$. Since we are employing a one-tailed test, the critical region consists of all values of $z = (\overline{X} - \mu_0)/\sigma_{\overline{X}} \geq 2.33$.

Therefore, the critical value of the sample statistic (the minimum value of $\overline{X}$ leading to rejection of H_0) is

$$\overline{X} = (2.33)\sigma_{\overline{X}} + \mu_0.$$

Thus the power equals the probability of obtaining this critical value in the distribution under H_1. The following steps are employed.

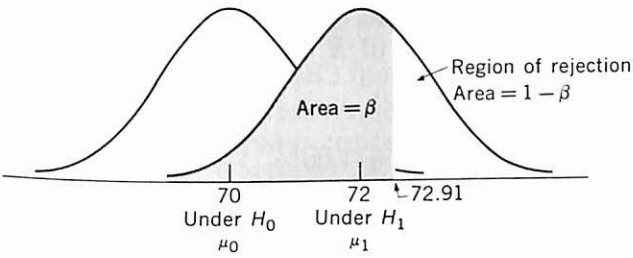


Figure 16.1
Region of rejection for H_0 in the distribution under H_1.

Step 1. Calculate the value of $\sigma_{\overline{X}}$:

$$\sigma_{\overline{X}} = \sigma/\sqrt{N} = 5/\sqrt{16} = 1.25.$$

Step 2. Determine the critical value of $\overline{X}$ ($\alpha = 0.01$, one-tailed test):

$$\overline{X} = (2.33)(1.25) + 70 = 72.91.$$

Step 3. Determine the probability of obtaining this critical value in the *true* sampling distribution under H_1. The critical value of $\overline{X}$ has a z-score, in the distribution under H_1, of

$$z = \frac{72.91 - \mu_1}{\sigma_{\overline{X}}} = \frac{72.91 - 72.00}{1.25} = 0.73.$$

Referring to Column C (Table A), we see that the probability of correctly rejecting H_0 is 23.27%. This probability is $1 - \beta$, or the power of the test. Incidentally, the probability of making a type II error (β) is 76.73%.

Figure 16.1 clarifies these relationships by indicating the region of rejection for H_0 in the *true* distribution under H_1, $1 - \beta$ or power. The shaded area indicates β, which is the probability of falsely accepting H_0.

16.3 THE EFFECT OF SAMPLE SIZE ON POWER

Power varies as a function of several different factors. Let us examine the effect of varying the size of the sample on the power of the test. For example, let us employ $N = 25$, in the problem described in the previous section, and see what effect this has on the power of the test.

Employing the same procedures as described above, we shall test the hypothesis:

$$H_0: \mu = \mu_0 = 70,$$

given that the true hypothesis is

$$H_1: \mu = \mu_1 = 72.$$

Since, with $\alpha = 0.01$ (one-tailed test), the critical region consists of all values of $z = (\overline{X} - \mu_0)/\sigma_{\overline{X}} \geq 2.33$, the critical value of $\overline{X} = (2.33)\sigma_{\overline{X}} + \mu_0$. Thus,

Step 1. The value of $\sigma_{\overline{X}}$ is

$$\sigma_{\overline{X}} = \sigma/\sqrt{N} = 5/\sqrt{25} = 1.00.$$

Step 2. The critical value of $\overline{X}$ is

$$\overline{X} = (2.33)(1.00) + 70 = 72.33.$$

Step 3. The critical value of $\overline{X}$ has a z-score in the distribution under H_1 of

$$z = \frac{72.33 - 72.00}{1.00} = 0.33.$$

Referring to Column C (Table A), we see that the power of the test is 37.07%.

We have seen that when $N = 16$, in our illustrative problem, then the power is 23.27%. When we increased our N to 25, power increased to 37.07%. Had we determined power for $N = 100$, for the above example, we would find that the power = 95.25%. Thus we may conclude that the power of a test is a function of N.

16.4 THE EFFECT OF α-LEVEL ON POWER

In our previous discussion on type I and type II errors (Section 11.6), we indicated that the lower we set α, the smaller the likelihood of a type I error, and the greater the likelihood of a type II error. Since β is the probability that a type II error will occur, and power $= 1 - \beta$, the higher the α-level chosen, the greater the power of the test. One can readily demonstrate this relationship between α and power by substituting a different α-level in the preceding problems and observing the change in power.

For example, employing $\alpha = 0.05$ (one-tailed test) with $N = 16$, we find that the critical region consists of all values of

$$z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} \geq 1.65.^*$$

Therefore the critical value of $\overline{X} = (1.65)\sigma_{\overline{X}} + \mu_0 = 72.06$. Thus power = 48.01% when $\alpha = 0.05$, as compared to 23.27% when $\alpha = 0.01$ in the example from Section 16.2.

---

* When $\alpha = 0.05$ (one-tailed test), the critical value of z is exactly halfway between 1.64 and 1.65. We shall employ $z = 1.65$ as the critical value so that $p < 0.05$ rather than $z = 1.64$, which results in $p > 0.05$.

16.5 THE EFFECT OF THE NATURE OF H_1 ON POWER

The power of a test is also a function of the nature of the alternative hypothesis. In the event that H_0 is actually false, the directional or one-tailed H_1 is more powerful than the two-tailed test so long as the parameter is in the predicted direction.

Table 16.1

Critical values of z required to reject H_0 at various α-levels as a function of the nature of H_1

Nature of H_1		
Directional (one-tailed test)		Nondirectional (two-tailed test)
$\alpha = 0.005$	$z = 2.58$	$z = \pm 2.81$
$\alpha = 0.01$	$z = 2.33$	$z = \pm 2.58$
$\alpha = 0.025$	$z = 1.96$	$z = \pm 2.24$
$\alpha = 0.05$	$z = 1.65$	$z = \pm 1.96$

Inspection of Table 16.1 reveals that the higher the α-level, the lower the absolute value of z required to reject H_0. We have already seen that power increases with increasing α. It follows that power increases as the critical value of z decreases. Table 16.1 shows that for any given α-level, the critical value of z is lower for a one-tailed test than for a two-tailed test. Therefore, an obtained z which is not significant for a two-tailed test may be significant for a one-tailed test. Thus the one-tailed test is more powerful than its two-tailed alternative, unless the parameter happens to lie in a direction opposite to the one predicted. In this case, the one-tailed test will be less powerful.

16.6 PARAMETRIC VS. NONPARAMETRIC TESTS: POWER

Another factor determining the power of a statistical test is the nature of the test itself. We can state as a general rule that for any given N, the parametric tests are more powerful than their nonparametric counterparts. It is primarily for this reason that we have deferred the discussion of statistical power until the present section of the text. For any given N, the parametric tests of significance (those assuming normally distributed populations with the same variance) entail less risk of a type II error. They are more likely to reject H_0 when H_0 is false. Thus given the choice between a nonparametric and a parametric test of significance, the parametric test should be employed so long as its underlying assumptions are fulfilled. However, as we shall see in the following chapters,

there are numerous situations in which the very nature of our data excludes the possibility of a parametric test of significance. We shall therefore be forced to employ less powerful nonparametric tests.*

Why do nonparametric tests have less power? Succinctly stated, the answer is that parametric statistical tests (as opposed to nonparametric tests) make maximum use of all the information that is inherent in the data when the populations are normally distributed. Let us look at a simple illustration. Imagine that we have obtained the following scores in the course of conducting a study: 50, 34, 21, 12, 10. Now, if we were to convert these scores into ranks (an operation basic to nonparametric statistics involving ordinal scales), we would obtain 1, 2, 3, 4, 5. Note that all the information concerning the *magnitudes* of the scores is lost when we convert to ranks. The difference between the scores of 50 and 34 becomes "equivalent" when expressed as ranks to the difference between, say, 12 and 10. This greater sensitivity of the parametric tests to the magnitudes of scores makes them a more accurate basis for arriving at probability values when the basic assumptions of cardinality are met.

16.7 CALCULATION OF POWER: TWO-SAMPLE CASE

So far, we have examined the effect of various factors on power, employing the one-sample case. All of the conclusions drawn apply equally to the two-sample case.

At this point, we should like to illustrate a sample problem in which we calculate the power of a test for the two-sample case.

Let us suppose that we have two populations with the following parameters:

$$\mu_1 = 80, \qquad \mu_2 = 75,$$

$$\sigma_1 = 6, \qquad \sigma_2 = 6.$$

If we draw a sample of nine cases from each of the two populations ($n_1 = 9$, $n_2 = 9$), we may test for the significance of the difference between the two sample means obtained. First, let us set up this problem in formal statistical terms.

1. *Null hypothesis* (H_0): The two samples were drawn from populations with equal means, that is, $\mu_1 = \mu_2$.
2. *Alternative hypothesis* (H_1): The two samples were drawn from populations with different means, that is, $\mu_1 \neq \mu_2$.

---

* It must be reiterated that the parametric tests are more powerful only when the assumptions underlying their use are valid. When the assumptions are not met, a nonparametric treatment may be as powerful as the parametric.

3. *Statistical test:* Since we are comparing two sample means drawn from normally distributed populations with known variances, z is the appropriate test statistic.
4. *Significance level:* $\alpha = 0.01$.
5. *Sampling distribution:* The sampling distribution of the statistic $(\overline{X}_1 - \overline{X}_2)$ is known to be a normal distribution.
6. *Critical region:* $|z_{0.01}| \geq 2.58$. Since H_1 is nondirectional, the critical region consists of all values of $z \geq 2.58$ and $z \leq -2.58$.

In other words, when

$$|z| = \left|\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{X}_1 - \overline{X}_2}}\right| \geq 2.58,$$

we will reject H_0. Since H_0 means that $\mu_1 - \mu_2 = 0$, the lower critical value of $(\overline{X}_1 - \overline{X}_2) = (-2.58)\sigma_{\overline{X}_1 - \overline{X}_2}$, and the upper critical value of $(\overline{X}_1 - \overline{X}_2) = (+2.58)\sigma_{\overline{X}_1 - \overline{X}_2}$.

Now, since we know H_0 to be false (that is, $\mu_1 - \mu_2 \neq 0$), the power of the test is equal to the probability of obtaining these critical values. Any obtained sample difference which is less than these critical values will lead to a type II error (i.e., acceptance of a false H_0).

We employ the following steps to calculate power:

Step 1. Calculate the value of $\sigma_{\overline{X}_1 - \overline{X}_2}$:

$$\sigma_{\overline{X}_1} = \frac{\sigma_1}{\sqrt{n_1}} = 2, \qquad \sigma_{\overline{X}_2} = \frac{\sigma_2}{\sqrt{n_2}} = 2,$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}} = 2.828.$$

Step 2. Determine the critical values of $(\overline{X}_1 - \overline{X}_2)$, $\alpha = 0.01$, two-tailed test. The lower critical value is $(\overline{X}_1 - \overline{X}_2) = (-2.58)(2.828) = -7.296$. The upper critical value is $(\overline{X}_1 - \overline{X}_2) = 7.296$.

Step 3. Determine the probability of obtaining these critical values in the *true* sampling distribution under H_1. The upper critical value of $(\overline{X}_1 - \overline{X}_2)$ has a z-score, in the distribution under H_1, of

$$z = \frac{7.296 - \mu_{\overline{X}_1 - \overline{X}_2}}{\sigma_{\overline{X}_1 - \overline{X}_2}} = \frac{7.296 - 5.0}{2.828} = 0.81.$$

Referring to Column C (Table A), we see that the area beyond a z of 0.81 is 20.90%. The lower critical value of $(\overline{X}_1 - \overline{X}_2)$ has a z-score of

$$z = \frac{-7.296 - 5.0}{2.828} = -4.35.$$

A z of -4.35 is so large that only a neglible proportion of area falls beyond it ($<0.003\%$). Thus the power $= 20.90\%$.

16.8 THE EFFECT OF CORRELATED MEASURES ON POWER

In Section 14.2 we indicated that when subjects have been successfully matched on a variable correlated with the criterion variable, a statistical test which takes this correlation into account provides a more powerful test than one which does not. This may be readily demonstrated.

Employing the data in the preceding problem, let us assume that the nine subjects drawn from population 1 are matched on a related variable with the nine subjects drawn from population 2 and that the correlation between these two variables is 0.80.

Since, with $\alpha = 0.01$ (two-tailed test), the critical region consists of all values of

$$|z| = \left|\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{X}_1-\overline{X}_2}}\right| \geq 2.58,$$

the critical values of $(\overline{X}_1 - \overline{X}_2) = (\pm 2.58)\sigma_{\overline{X}_1-\overline{X}_2}$.

Step 1. The value of $\sigma_{\overline{X}_1-\overline{X}_2}$ is

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} - 2r\sigma_{\overline{X}_1}\sigma_{\overline{X}_2}} = 1.26.$$

Step 2. The critical values of $(\overline{X}_1 - \overline{X}_2)$ are

$$(\overline{X}_1 - \overline{X}_2) = (\pm 2.58)(1.26) = \pm 3.25.$$

Step 3. The lower critical value of $(\overline{X}_1 - \overline{X}_2)$ has a z-score, in the distribution under H_1, of

$$z = \frac{(-3.25 - 5.00)}{1.26} = -6.55.$$

Referring to Table A, we find that the area beyond a z of -6.55 is negligible. Therefore, power will be determined according to the upper critical value.

The upper critical value of $(\overline{X}_1 - \overline{X}_2)$ has a z of

$$z = \frac{3.25 - 5.00}{1.26} = -1.39.$$

To find the probability of obtaining $(\overline{X}_1 - \overline{X}_2) \geq 3.25$, we refer to Table A to find the area *above* a z of -1.39. Thus power $= 100.00\% - 8.23\% = 91.77\%$. Since power $= 1 - \beta$, the probability of a type II error (β) is 8.23%.

16.9 POWER, TYPE I AND TYPE II ERRORS

Let us take a moment to tie together some of our observations about power and type I and type II errors. To begin with, we must emphasize the fact that there are only two possibilities with respect to the null hypothesis, i.e., either it is true (for example, $\mu_1 = \mu_2$) or it is not true (for example, $\mu_1 \neq \mu_2$). These are two mutually exclusive situations. Now, since a type I error is defined as the probability of rejecting H_0 when it is true, two points should immediately be obvious.

1. If H_0 is *false*, the probability of a type I error is zero.

2. It is only when we *reject* H_0 that any possibility exists for a type I error. Such an error will be made only when H_0 is true, in which case, the probability of a type I error is α.

Further, since a type II error is defined as the probability of accepting H_0 when it is false, we arrive at the following conclusions.

1. If H_0 is *true*, the probability of a type II error is zero.

2. It is only when we *accept* H_0 that any possibility exists for a type II error. Such an error will be made only when H_0 is false, in which case the probability of a type II error is β. It should be clear, then, that the concept of power, which is defined in terms of a type II error $(1 - \beta)$, applies only when H_0 is not true.

Table 16.2 summarizes the probabilities associated with acceptance or rejection of H_0 depending on the true state of affairs.

Table 16.2
True Status of H_0

		H_0 true	H_0 false
Decision	Accept H_0	Correct $1 - \alpha$	Type II error β
	Reject H_0	Type I error α	Correct $1 - \beta$

16.10 POWER EFFICIENCY OF A STATISTICAL TEST

In Section 16.6 we pointed out that when the underlying assumptions can be considered valid, parametric tests are more powerful for any given N than nonparametric tests. However, it is also true that when nonparametric tests are to be utilized, we can make any specific nonparametric test as powerful as a

parametric test by employing a larger sample size. Thus test A may be more powerful than test B when the N's are equal, but B may be as powerful as A when an N of, say, 40 is used compared to an N of 30 with test A.

The concept of *power efficiency* is concerned with the increase in sample size required to make one test as powerful as a competing test. Let us assume that test A is the most powerful for the type of data which we are analyzing. Let us assume that test B is equal in power to test A when their N's are 40 and 30 respectively. We shall let N_b represent the N required to make it as powerful as test A when N_a is used. The power efficiency of test B may now be stated:

$$\text{Power efficiency of test } B = 100\,\frac{N_a}{N_b}\text{ percent.} \tag{16.2}$$

Thus in the above example, the power-efficiency of test B relative to test A is $100(\frac{30}{40})$ or 75%. Therefore, assuming that all the assumptions for employing test A are met, we shall have to use four cases of test B for every three cases of test A to achieve equal power. Of course, if the assumptions underlying test A are not met, the concept of power efficiency has no meaning since test A should not be employed.

CHAPTER SUMMARY

In this chapter, we discussed two important concepts: power and power efficiency. Power is defined as the probability of rejecting H_0 when it is actually false, i.e., *power* $= 1 - \beta$.

We demonstrated the calculation of power for the one-sample and the two-sample cases when H_0 is known to be false and the true value of the parameter under H_1 is known.

The calculation of power requires that we compute

1. the standard error of the sampling distribution under both H_0 and H_1,
2. the critical value of the sample statistic [$\overline{X}$ in the one-sample case; $(\overline{X}_1 - \overline{X}_2)$ in the two-sample case],
3. the probability of obtaining this critical value in the sampling distribution under H_1. This probability is the power of the test.

We showed that power varies as a function of:

1) sample size,
2) α level,
3) the nature of H_1,
4) the nature of the statistical test,
5) the use of correlated measures.

Power efficiency is concerned with the increase in sample size of a given test necessary to make it as powerful as another test employing a smaller N. Symbolically, the power efficiency of test B relative to test A may be represented as

$$\text{power efficiency of test } B = (100)\,\frac{N_a}{N_b}\text{ percent.}$$

Terms to Remember:

Power
Power-efficiency

EXERCISES

1. Test A has a power efficiency of 80% relative to test B. If, in test B we employed a total of 24 subjects, what is the N required to achieve equal power with test A?
2. Employing the sample problem in Section 16.3, calculate power when $N = 100$.
3. Employing the sample problem in Section 16.7, demonstrate the effect of the nature of H_1 on power by calculating the power when $\alpha = 0.01$, *one*-tailed test, that is, $H_1: \mu_1 > \mu_2$.
4. Given two normal populations:

$$\mu_1 = 100, \qquad \mu_2 = 90,$$
$$\sigma_1 = 10, \qquad \sigma_2 = 10.$$

   Employ $\alpha = 0.01$, two-tailed test.
   a) If a sample of 25 cases is drawn from each population, find
      1) the probability of a type I error,
      2) the probability of a type II error,
      3) the power of the test.
   b) If two samples of 25 cases each is drawn from Population 1, find
      1) the probability of a type I error,
      2) the probability of a type II error,
      3) the power of the test.
5. For Problem 6, Chapter 13, calculate the power for each of the four examples, employing $\alpha = 0.01$, one-tailed test. Which of the factors influencing power do these examples illustrate?
6. In Chapter 14, Problems 3 through 5, we saw that the use of correlated samples produced a t-ratio farther removed from the region of rejection than the t-ratio based on independent samples. This would appear to contradict Section 16.8. Reconcile this disparity.

Statistical Inference With Categorical Variables

17

17.1 INTRODUCTION

In recent years there has been a broadening in both the scope and the penetration of research in the behavioral and social sciences. Much provocative and stimulating research has been initiated in such diverse areas as personality, psychotherapy, group processes, economic forecasts, etc. New variables have been added to the arsenal of the researcher, many of which do not lend themselves to traditional parametric statistical treatment, either because of the scales of measurement employed or because of flagrant violations of the assumptions of these parametric tests. For these reasons, many new statistical techniques have been developed.

Parametric techniques are usually preferable because of their greater sensitivity. This generalization is not true, however, when the underlying assumptions are seriously violated. Indeed, under certain circumstances (e.g., badly skewed distributions, particularly with small n's), a nonparametric test may well be as powerful as its parametric counterpart.* Consequently, the researcher is frequently faced with the difficult choice of a statistical test appropriate for his data.

At this point, let us interject a word of caution with respect to the choice of statistical tests of inference. For heuristic purposes, we will take a few sample problems and subject them to statistical analyses employing several different tests of significance. This procedure will serve to clarify points of differences among the various tests. However, you may inadvertently draw an erroneous conclusion, namely, that the researcher first collects his data, and then "shops

* Numerous investigators have demonstrated the robustness of the t and F tests; i.e., even substantial departures from the assumptions underlying parametric tests do not seriously affect the validity of statistical inferences. For articles dealing with this topic, see Runyon, Haber, and Badia, *Readings in Statistics*, Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1970.

around" for a statistical test which provides the most sensitive test for any differences which exist. Actually, nothing could be further from the truth. The null hypothesis, alternative hypothesis, statistical test, sampling distribution, and level of significance should all be specified in *advance* of the collection of data. If one "shops around," so to speak, after the collection of the data, he tends to maximize the effects of any chance differences which favor one test over another. As a result, the possibility of a type I error (rejecting the null hypothesis when it is true) is substantially increased.

Usually, we do not have the problem of choosing statistical tests in relation to categorical variables because nonparametric tests alone are suitable for enumerative data.

The problem of choosing a statistical treatment usually arises when we employ small samples* and/or when there is doubt concerning the normality of the underlying population distribution. In Chapter 18 we shall demonstrate several nonparametric statistical tests employed when such doubts arise.

17.2 THE BINOMIAL TEST

In Section 2.4, when discussing various scales of measurement, we pointed out that the observation of unordered variables constitutes the lowest level of measurement. In turn, the simplest form of nominal scale is one which contains only two classes or categories, and is referred to variously as a two-category or dichotomous population. Examples of two-category populations are numerous, e.g., male and female, right and wrong on a test item, married and single, juvenile delinquent and nondelinquent, literate and illiterate. Some of these populations may be thought of as inherently dichotomous (e.g., male vs. female) and therefore not subject to measurement on a higher-level numerical scale, whereas others (e.g., literate and illiterate) may be thought of as continuous, varying from the absence of the quality under observation to different degrees of its manifestation. Obviously, whenever possible, the data we collect should be at the highest level of measurement that we can achieve. However, for a variety of reasons, we cannot always scale a variable at the ordinal level or higher. Nevertheless, we are called upon to collect nominally scaled data and to draw inferences from these data.

In a two-category population, we define P as the proportion of cases in one class and $Q = 1 - P$ as the proportion in the other class.

In Chapter 11 we demonstrated the use of the binomial sampling distribution to test hypotheses concerning the value of P under H_0. For illustrative purposes

---

* When large samples are employed, the parametric tests are almost always appropriate because of the *central-limit theorem* (Section 12.2).

we restricted our discussion to H_0: $P = Q = \frac{1}{2}$. However, it is possible to test hypotheses concerning *any* value of P.

The probabilities associated with specific outcomes may be obtained by employing formula (17.1):

$$p(x) = \frac{N!}{x!(N-x)!} P^x Q^{N-x}, \tag{17.1}$$

where

$x =$ the number of objects in one category or the number of successes,

$N - x =$ the number of objects in the remaining category or the number of failures,

$N =$ the total number of objects or total number of trials,

$p(x) =$ the probability of x objects in one category,

! the factorial sign directs us to multiply the indicated value by all integers less than it but greater than zero, e.g., if $N = 5$, $N! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$.*

A sample problem should serve to illustrate the use of formula (17.1) in the testing of hypotheses.

17.2.1 Sample Problem

The dean of students in a large university claims that ever since the sale of cigarettes was prohibited on campus, the proportion of students who smoke has dropped to 0.30. However, previous observations at other institutions, where the sale of cigarettes has been banned, have found little effect on smoking behavior; i.e., far greater than 0.30 of the students continue to smoke.

Nine students, selected at random, are asked to indicate whether or not they smoke. Six of these students respond in the affirmative.

To test the validity of the dean's claim, we will let P represent the proportion of students in the population who smoke, and Q, the proportion of students who do not smoke.

1. *Null hypothesis* (H_0)*:* $P = 0.30$, $Q = 0.70$.
2. *Alternative hypothesis* (H_1)*:* $P > 0.30$, $Q < 0.70$. Note that H_1 is directional.
3. *Statistical test:* Since we are dealing with a two-category population, the binomial test is appropriate.

---

* In obtaining the binomial probabilities, it is important to remember that $0! = 1$ and any value other than zero raised to the zero power equals 1, i.e., $X^0 = 1$.

4. *Significance level:* $\alpha = 0.05$.
5. *Sampling distribution:* The sampling distribution is given by the binomial expansion, formula (11.1).
6. *Critical region:* The critical region consists of all values of x which are so large that the probability of their occurrence under H_0 is less than or equal to 0.05. Since H_1 is directional, the critical region is one-tailed.

In order to determine whether the obtained x lies in the critical region, we must obtain the sum of the probabilities associated with $x = 9$, $x = 8$, $x = 7$, and $x = 6$.

To illustrate, employing formula (17.1), we find that the probability of $x = 9$, that is, if $P = 0.3$ (under H_0) the probability that all of the 9 students smoke, is

$$\begin{aligned} p(9) &= \frac{9!}{9!(9-9)!}(0.3)^9(0.7)^{9-9} \\ &= (0.3)^9 = 0.000019683. \end{aligned}$$

The probability of $x = 8$ is

$$p(8) = 9(0.3)^8(0.7)^1 = 0.000413343.$$

The probability of $x = 7$ is

$$p(7) = 36(0.3)^7(0.7)^2 = 0.003857868.$$

Finally, the probability of $x = 6$ is

$$p(6) = 84(0.3)^6(0.7)^3 = 0.021003948.$$

Thus the probability that at least 6 out of 9 students smoke when $P = 0.30$ is the sum of the above probabilities, that is, $p(x \geq 6) = 0.025$.

Decision: Since the obtained probability is less than 0.05, we may reject H_0.

The above example was offered to illustrate the application of formula (17.1) when $P \neq Q \neq \frac{1}{2}$. However, when $N \leq 10$, and $P \neq Q \neq \frac{1}{2}$, Table N may be employed to obtain the critical values of x directly for selected values of P and Q. Referring to Table N, for $N = 9$ and $P = 0.30$, we find that x must be equal to or greater than 6 to reject H_0 at $\alpha = 0.05$.

Incidentally, when $P = Q = \frac{1}{2}$ and $N \leq 25$, Table M provides one-tailed probability values of x when x is defined as the smaller of the observed frequencies. For example, if we toss a coin ten times ($N = 10$), we find that the probability of obtaining two or fewer heads ($x \leq 2$) is 0.055.

17.3 NORMAL CURVE APPROXIMATION TO BINOMIAL VALUES

As N increases, the binomial distribution approaches the normal distribution. The approximation is more rapid as P and Q approach $\frac{1}{2}$. On the other hand, as P or Q approach zero, the approximation to the normal curve becomes poorer for any given N. A good rule of thumb to follow, when considering the normal curve approximation to the binomial, is that the product NPQ should equal at least 9 when P approaches 0 or 1. Accepting these restrictions, we see that the sampling distribution of x, defined as the number of objects in one category, is normal with a mean equal to NP and a standard deviation equal to $\sqrt{NPQ}$.

To test the null hypothesis, we put x into standarized form:

$$z = \frac{x - NP}{\sqrt{NPQ}}. \tag{17.2}$$

The distribution of the z-scores is approximately normal with a mean equal to zero and a standard deviation equal to one. Thus the probability of any given x equals the probability of its corresponding z-score. However, the normal distribution is based on a continuous variable whereas the binomial distribution is discontinuous. The approximation to the normal distribution becomes better if a correction for continuity is made. This correction consists of subtracting the constant 0.5 from the absolute value of $x - NP$. Thus z becomes

$$z = \frac{|x - NP| - 0.5}{\sqrt{NPQ}}. \tag{17.3}$$

Let us look at an example in which $x = 5$, $N = 20$, and $P = Q = \frac{1}{2}$. To obtain the probability of $x \leq 5$, we put x into standardized form, and employing the correction for continuity, we have

$$z = \frac{|5 - 10| - 0.5}{\sqrt{(20)(0.5)(0.5)}}$$

$$= \frac{4.50}{\sqrt{5.00}} = \frac{4.50}{2.24} = 2.01.$$

Locating this value in Column C (Table A), we find a probability value of 0.0222. Note that if we had looked up the result in Table M under $x = 5$ and $N = 20$, we would obtain a one-tailed p-value of 0.021. In other words, the approximation is extremely good. To obtain the two-tailed p-value we would, of course, double the obtained p. Thus in the current example, $p = 0.04$.

17.4 THE χ^2 ONE-VARIABLE CASE

Let us suppose that you are a market researcher hired by a soap manufacturer to conduct research on the packaging of his product. Realizing that color may be an important determinant of consumer selection, you conduct the following study. Six hundred housewives, selected by an accepted sampling technique, are each given three differently packaged cakes of the same brand of soap. They are told that the three soaps are made according to different formulas and that the distinctive coloring of the packages is merely to aid their identification of each soap. One month later, you inform each housewife that she is to receive a free case of soap of her own choosing. Their selections are listed below.

Color of Wrapper			
	Red	White	Brown
Number of housewives selecting	200	300	100

This is the type of problem for which the χ^{2*} one-variable test is ideally suited. In single-variable applications, the χ^2 test has been described as a "goodness of fit" technique: it permits us to determine whether or not a significant difference exists between the *observed* number of cases falling into each category, and the *expected* number of cases, based on the null hypothesis. In other words, it permits us to answer the question, "How well does our observed distribution fit the theoretical distribution?"

What we require, then, is a null hypothesis which allows us to specify the frequencies that would be expected in each category and, secondly, a test of this null hypothesis. The null hypothesis may be tested by

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}, \tag{17.4}$$

where

$f_o =$ the observed number in a given category,

$f_e =$ the expected number in that category,

$\sum_{i=1}^{k}$ directs us to sum this ratio over all k categories.

* The symbol χ^2 will be used to denote the test of significance as well as the quantity obtained from applying the test to observed frequencies whereas the word, "chi-square," will refer to the theoretical chi-square distribution.

As is readily apparent, if there is close agreement between the observed frequencies and the expected frequencies, the resulting χ^2 will be small, leading to a failure to reject the null hypothesis. As the discrepancy $(f_o - f_e)$ increases, the value of χ^2 increases. The larger the χ^2, the more likely we are to reject the null hypothesis.

In the above example, the null hypothesis would be that there is an equal preference for each color, i.e., 200 is the expected frequency in each category. Thus

$$\begin{aligned}\chi^2 &= (200 - 200)^2/200 + (300 - 200)^2/200 + (100 - 200)^2/200 \\ &= 0 + 50.00 + 50.00 \\ &= 100.00.\end{aligned}$$

In studying the Student t-ratio (Section 12.5) we saw that the sampling distributions of t varied as a function of degrees of freedom. The same is true for χ^2. However, assignment of degrees of freedom with the Student t-ratio are based on N, whereas, for χ^2, the degrees of freedom are a function of the number of categories (k). In the one-variable case, $\text{df} = k - 1$.* Table B lists the critical values of χ^2 for various α-levels. If the obtained χ^2 value exceeds the critical value at a given probability level, the null hypothesis may be rejected at that level of significance.

In the above example, $k = 3$. Therefore, $\text{df} = 2$. Employing $\alpha = 0.01$, we find that Table B indicates that a χ^2 value of 9.21 or greater is required for significance. Since our obtained value of 100.00 is greater than 9.21, we may reject the null hypothesis and assert, instead, that color is a significant determinant for soap preference among women.

17.5 THE χ^2 TEST OF THE INDEPENDENCE OF CATEGORICAL VARIABLES

So far in this chapter, we have been concerned with the one-variable case. In practice, employing categorical variables, we do not encounter the one-variable case too frequently. More often, we ask questions concerning the interrelationships between and among variables. For example, we may ask:

Is there a difference in the crime rate of children coming from different socioeconomic backgrounds?

Is there a difference in the recovery rates of patients undergoing various forms of psychotherapeutic treatment?

If we are conducting an opinion poll, can we determine whether there is a difference between males and females in their opinions about a given issue?

These are but a few examples of problems for which the χ^2 technique has an application. You could undoubtedly extend this list to include many campus

---

* Since the marginal total is fixed, only $k - 1$ categories are free to vary.

activities, such as attitudes of fraternity brothers vs. nonfraternity students toward certain basic issues (e.g., cheating on exams), differences in grading practices among professors in various departments of study. All the above problems have some things in common: (1) They deal with two or more nominal categories in which (2) the data consist of a frequency count which is tabulated and placed in the appropriate cells.

Table 17.1

2×2 contingency table showing the number of men and women indicating characteristics of automobiles they like best

	Response to question		
Sex of respondents	Appearance	Performance	Row marginal
Male	(a) 75	(b) 125	200
Female	(c) 150	(d) 100	250
Column marginal	225	225	450

These examples also share an additional and more important characteristic: (3) There is no immediately obvious way to assign expected frequency values to each category. However, as we shall point out, shortly, what we must do is base our expected frequencies on the obtained frequencies themselves.

Let us take a look at a hypothetical example. It is frequently claimed that men select cars primarily by performance characteristics, whereas women select cars primarily by appearance characteristics. Two hundred men and 250 women were asked the question, "What is the one characteristic of your present automobile which is most satisfying to you?" The responses to the question were placed in one of two subgroups, depending on whether performance characteristics or appearance was mentioned. The results are found in Table 17.1. Inspection of the table reveals that the results confirm the claim.

We must now apply a test of significance. In formal statistical terms,

1. *Null hypothesis* (H_0)*:* There is no difference between men and women in their preferences concerning "most liked" automobile characteristics.
2. *Alternative hypothesis* (H_1)*:* There is a difference between men and women in their preferences concerning "most liked" characteristic.
3. *Statistical test:* Since the two groups (male and female) are independent, and the data are in terms of frequencies in discrete categories, the χ^2 test of independence is the appropriate statistical test.

4. *Significance level:* $\alpha = 0.05$.
5. *Sampling distribution:* The sampling distribution is the chi-square distribution with df $= (r - 1)(c - 1)$.

   Since marginal totals are fixed, the frequency of only one cell is free to vary. Therefore, we have a one-degree-of-freedom situation. The general rule for finding df in the two-variable case is $(r - 1)$ $(c - 1)$, in which $r =$ number of rows and $c =$ number of columns. Thus in the present example, df $= (2 - 1)(2 - 1) = 1$.
6. *Critical region:* Table B (in Appendix III) shows that for df $= 1$, $\alpha = 0.05$, the critical region consists of all values of $\chi^2 \geq 3.84$. χ^2 is calculated from the formula

$$\chi^2 = \sum_{r=1}^{r} \sum_{c=1}^{c} \frac{(f_o - f_e)^2}{f_e}, \tag{17.5}$$

   where

   $\sum_{r=1}^{r} \sum_{c=1}^{c}$ directs us to sum this ratio over both rows and columns.

The *main problem* now is to decide on a basis for determining the expected cell frequencies. Let us concentrate for a moment on cell (a). The two marginal totals common to cell (a) are row 1 marginal and column 1 marginal. If the null hypothesis is correct, we would expect the same proportion of men and of women to cite appearance as a deciding factor in automobile preference. Since 200 of the total sample of 450 are men, we would expect that $\frac{200}{450} \times 225$ men would be found in cell (a). This figure comes to 100. Since we have a one-degree-of-freedom situation, the expected frequencies in all the remaining cells are determined as soon as we have calculated one expected cell frequency. Consequently, we may obtain all the remaining expected cell frequencies by subtracting from the appropriate marginal totals. Thus cell (b) is $200 - 100$ or 100; cell (d) is $225 - 100$ or 125; and cell (c) is $225 - 100$ or 125. To make certain that no error was made in our original calculation of the expected frequency for cell (a), it is wise to independently calculate the expected frequency for *any* of the remaining cells. If this figure agrees with the result we obtained by subtraction, we may feel confident that we made no error. Thus the expected frequency for cell (d), obtained through direct calculation, is

$$\tfrac{250}{450} \times 225 = 125.$$

Since this figure does agree with the result obtained by subtraction, we may proceed with our calculation of the χ^2 value.

Table 17.2

Number of men and women indicating characteristics of automobiles they prefer most (expected frequencies within parentheses)

Sex of respondents	Response to question: Appearance	Performance	Row marginal
Male	75 (100)	125 (100)	200
Female	150 (125)	100 (125)	250

Incidentally, you may have noted that there is a simple rule which may be followed in determining expected frequency of a given cell; you multiply the marginal frequencies common to that cell and divide by N.

Table 17.2 presents the obtained data, with the expected cell frequencies in the lower right-hand corner of each cell.

Now, all that remains to calculate is the χ^2 value. However, as in the normal approximation to the binomial, the empirical distributions of categorical variables are discrete, whereas the theoretical distributions of chi-square are continuous. Consequently, in the one-degree-of-freedom situation, a correction for continuity is required to obtain a closer approximation of the obtained χ^2 values to the theoretical distribution. This correction consists of subtracting 0.5 from the absolute difference $|f_o - f_e|$. Thus, in the one-degree-of-freedom situation, the formula for calculating χ^2 becomes

$$\chi^2 = \sum_{r=1}^{2} \sum_{c=1}^{2} \frac{(|f_o - f_e| - 0.5)^2}{f_e}, \tag{17.6}$$

where

$$\chi^2 = \frac{(|75 - 100| - 0.5)^2}{100} + \frac{(|125 - 100| - 0.5)^2}{100} + \frac{(|150 - 125| - 0.5)^2}{125} + \frac{(|100 - 125| - 0.5)^2}{125}$$

$$= \frac{600.25}{100} + \frac{600.25}{100} + \frac{600.25}{125} + \frac{600.25}{125} = 21.60.$$

Since the χ^2 value 21.60 is greater than 3.84, required for significance at the 0.05 level, we may reject H_0. In other words, we may conclude that men and women show a differential basis for car preference and, more specifically, women

state their preference on the basis of appearance characteristics more often than men.

In research, we often find that we have more than two subgroups within a nominal class. For example, we might have three categories in one scale and four in another, resulting in a 3×4 contingency table. The procedure for obtaining the expected frequencies is the same as the one for the 2×2 contingency table. Of course, the degrees of freedom will be greater than 1 (e.g., 3×4 contingency table, df $= 6$). Thus no correction for continuity is necessary.

17.6 LIMITATIONS IN THE USE OF χ^2

A fundamental assumption in the use of χ^2 is that each observation or frequency is independent of all other observations. Consequently, one may not make several observations on the same individual and treat each as though it were independent of all the other observations. Such an error produces what is referred to as an *inflated* N, that is, you are treating the data as though you had a greater number of independent observations than you actually have. This error is extremely serious and may easily lead to the rejection of the null hypothesis when it is, in fact, true.

Consider the following hypothetical example. Imagine that you are a student in a sociology course and, as a class project, you decide to poll the student body to determine whether male and female students differ in their opinions on some issue of contemporary significance. Each of 15 members of the class is asked to obtain replies from 10 respondents, 5 male and 5 female. The results are listed in Table 17.3.

Employing $\alpha = 0.05$, we find that the critical region consists of all the values of $\chi^2 \geq 3.84$. Since the obtained χ^2 of $9.67 > 3.84$, you reject the null hypothesis of no difference in the opinions of male and female students on the issue in question. You conclude, instead, that approval of the issue is dependent on the sex of the respondent.

Table 17.3

Sex	Response to question: Approve	Disapprove	
Male	30 (40)	45 (35)	75
Female	50 (40)	25 (35)	75
	80	70	150
	$\chi^2 = 9.67$		

Table 17.4

	Response to question		
Sex	Approve	Disapprove	
Male	28 (32.5)	37 (32.5)	65
Female	32 (27.5)	23 (27.5)	55
	60	60	120
	$\chi^2 = 2.15$		

Subsequent to the study, you discover that a number of students were inadvertently polled as many as two or three times by different members of the class. Consequently, the frequencies within the cells are not independent since some individuals had contributed as many as two or three responses. In a reanalysis of the data, in which only one frequency per respondent was permitted, we obtained the results shown in Table 17.4.

Note that now the obtained χ^2 of $2.15 < 3.84$; thus you must accept H_0. The failure to achieve independence of responses resulted in a serious error in the original conclusion. Incidentally, you should note that the requirement of independence within a cell or condition is basic to *all* statistical tests. We have mentioned this specifically in connection with the χ^2 test because violations may be very subtle and not easily recognized.

When discussing the normal approximation to the binomial, we pointed out that the extent of the approximation becomes less as P and Q diverge from $\frac{1}{2}$. We stated, as a rule of thumb, that NPQ should equal at least 9 to justify the normal probability model. A similar stricture applies to the χ^2 test. With small N's or when the expected proportion in any cell is small, the approximation of the sample statistics to the chi-square distribution may not be very close. A rule which has been generally adopted, in the one-degree-of-freedom situation, is that the expected frequency in all cells should be equal to or greater than 5. When df > 1, the expected frequency should be equal to or greater than 5 in at least 80% of the cells. When these requirements are not met, other statistical tests are available. (See Siegel, 1956.)

CHAPTER SUMMARY

In this chapter we have discussed four tests of significance employed with categorical variables, i.e., the binomial test, the normal approximation to the binomial, the χ^2 one-variable test and the χ^2 two-variable test.

1. We saw that the binomial may be employed to test null hypotheses when frequency counts are distributed between two categories or cells. When $P = Q = \frac{1}{2}$ and when $N \leq 25$, Table M provides the one-tailed probabilities directly. When $P \neq Q$, probabilities may be obtained by employing formula (17–1). which is based on the binomial expansion. When $N \leq 10$, Table N provides the critical values for the selected values of P and Q. A rule of thumb is that NPQ must equal or exceed 9 as P approaches 0 or 1 to permit the use of the normal approximation to the binomial.

2. The χ^2 one-variable test has been described as a "goodness of fit" technique, permitting us to determine whether or not a significant difference exists between the *observed* number of cases appearing in each category and the *expected* number of cases specified under the null hypothesis.

3. The χ^2 two-variable case may be employed to determine whether two variables are related or independent. If the χ^2 value is significant, we may conclude that the variables are interdependent, or related. In this chapter, we restricted our discussion to the 2×2 contingency table. However, the procedures are easily extended to include more than two categories within each variable.

4. Finally, we discussed two limitations on the use of the χ^2 test. In the one-degree-of-freedom situation, the expected frequency should equal or exceed 5 to permit the use of the χ^2 test. When df > 1, the expected frequency in 80% of the cells should equal or exceed 5. A second, and most important, restriction is that the frequency counts must be independent of one another. Failure to meet this requirement results in an error known as the *inflated* N and may well lead to the rejection of the null hypothesis when it is true (Type I error).

Terms to Remember:

Two category (dichotomous) populations
Binomial test
Binomial expansion
Normal approximation to binomial
Correction for continuity
χ^2 *one-variable case*
"Goodness of fit" test
χ^2 *test of independence*
"Inflated N"

EXERCISES

1. In 9 tosses of a single coin,
   a) what is the probability of obtaining exactly 7 heads?
   b) what is the probability of obtaining as many as 8 heads?
   c) what is the probability of obtaining a result as rare as 8 heads?
2. A revelation to many students is the surprisingly low probability of obtaining a passing grade on multiple choice examinations when the selection of alternatives is made purely on a chance basis (i.e., without any knowledge of the material covered on the exam).

If 60% is considered a passing grade on a 40-item multiple choice test, determine the probability of passing when there are
a) four alternatives, b) three alternatives, c) two alternatives.

3. In reference to Problem 4, Chapter 15, show how the binomial expansion might be employed to determine the probability of one success (i.e., rejection of H_0 at $\alpha = 0.05$) out of ten attempts.

4. A study was conducted in which three groups of rats (5 per group) were reinforced under three different schedules of reinforcement (100% reinforcement, 50% reinforcement, 25% reinforcement). The number of bar-pressing responses obtained during extinction are shown below.

100%	50%	25%
615	843	545

Criticize the use of chi-square as the appropriate statistical technique.

5. The World Series may last from four to seven games. During the period 1922–1965, the distribution of the number of games played per series was as follows:

Number of games	4	5	6	7
Frequency of occurrence	9	8	9	18

For these data, test the hypothesis that each number of games is equally likely to occur.

6. In a large eastern university, a study of the composition of the student council reveals that 6 of its 8 members are political science majors. In the entire student body of 1200 students, 400 are political science majors. Set this study up in formal statistical terms and draw the appropriate conclusions.

7. A study was conducted to determine if there is a relationship between socio-economic class and attitudes toward a new urban-renewal program. The results are listed below.

		Disapprove	Approve
Socio-economic class	Middle	90	60
	Lower	200	100

Set this study up in formal statistical terms and draw the appropriate conclusion.

8. Construct the sampling distribution of the binomial when $P = 0.20$, $Q = 0.80$, and $n = 6$.

9. Out of 300 castings on a given mold, 27 were found to be defective. Another mold produced 31 defective castings in 500. Determine whether there is a significant difference in the proportion of defective castings produced by the two molds.

10. Employ the χ^2 test, one-variable case, for the example shown in Section 17.3 of the text. Verify that $\chi^2 = z^2$ in the one-degree-of-freedom situation.

11. In a study concerned with preferences of packaging for potato chips, 100 women in a high income group and 200 women in a lower income group were interviewed. The results of their choices follow:

	Upper income group	Lower income group
Prefer metallic package	36	84
Prefer waxed paper package	39	51
Prefer cellophane package	16	44
Have no preference	9	21

What conclusions would you draw from these data?

12. Suppose that 100 random drawings from a deck of cards produced 28 hearts, 19 clubs, 31 diamonds, and 22 spades. Would you consider these results unusual?

13. In polling 46 interviewees drawn at random from a specified population, we find that 28 favor and 18 oppose a certain routing of a highway. Test the hypothesis that the sample was drawn from a population in which $P = Q = \frac{1}{2}$. Use the normal approximation to the binomial and the χ^2 one-variable case. Verify that $z = \sqrt{\chi^2}$.

Statistical Inference With Ordinally Scaled Variables

18

18.1 INTRODUCTION

In the previous chapter, we pointed out that the researcher is frequently faced with a choice as to which statistical test is appropriate for his data. You will recall that this was not really a problem in relation to categorical variables because nonparametric tests alone are suitable for nominally scaled data.

In this chapter we shall discuss several statistical techniques which are frequently employed as alternatives to parametric tests.

18.2 MANN-WHITNEY *U*-TEST

The Mann-Whitney U-test is one of the most powerful nonparametric statistical tests, since, as we shall see, it utilizes most of the quantitative information that is inherent in the data. It is most commonly employed as an alternative to the Student t-ratio when the measurements fail to achieve interval scaling or when the researcher wishes to avoid the assumptions of the parametric counterpart.

Imagine that we have drawn two independent samples of n_1 and n_2 observations. The null hypothesis is that both samples are drawn from populations with the same distributions. The two-tailed alternative hypothesis, against which we test the null hypothesis, is that the parent populations, from which the samples were drawn, are different. Imagine, further, that we combine the $n_1 + n_2$ observations and assign a rank of 1 to the smallest value, a rank of 2 to the next smallest value, and continue until we have assigned ranks to all the observations. Let us refer to our two groups as E and C respectively. If we were to count the number of times each C precedes each E in the ranks, we would expect, under the null hypothesis, that it would equal the number of times each E precedes a C. In other words, if there is no difference between the two groups,

the order of E's preceding C's, and vice versa, should be random. However, if the null hypothesis is not true, we would expect a bulk of the E-scores or the C-scores to precede their opposite number.

Let us take an example. Suppose you have the hypothesis that leadership is a trainable quality. You set up two groups, one to receive special training in leadership (E) and the other to receive no special instruction (C). Following the training, independent estimates of the leadership qualities of all the subjects are obtained. The results are:

E-scores	12	18	31	45	47
C-scores	2	8	15	19	38

In employing the Mann-Whitney test, we are concerned with the sampling distribution of the statistic "U." To find U, we must first rank all the scores from the lowest to the highest, retaining the identity of each score as E or C (Table 18.1).

Table 18.1

Rank	1	2	3	4	5	6	7	8	9	10
Score	2	8	12	15	18	19	31	38	45	47
Condition	C	C	E	C	E	C	E	C	E	E

You note that the number of E's preceding C's is less than the number of C's preceding E's. The next step is to count the number of times each E precedes a C. Note that the first E (score of 12) precedes three C's (scores of 15, 19, and 38 respectively). The second E (score of 18) precedes two C's (scores of 19 and 38). The third E (score of 31) precedes one C (score of 38). Finally, the last two E's precede no C's. U is the sum of the number of times each E precedes a C. Thus in our hypothetical problem, $U = 3 + 2 + 1 + 0 + 0 = 6$. Had we concentrated on the number of times C's precede E's we would have obtained a sum of $5 + 5 + 4 + 3 + 2 = 19$. We shall refer to this greater sum as U'. Under the null hypotheses, U and U' should be equal. The question is whether the magnitude of the observed difference is sufficient to warrant the rejection of the null hypothesis.

The sampling distribution of U under the null hypothesis is known. Tables I_1 through I_4 show the values of U and U' which are significant at various α levels. To be significant at a given α-level, the obtained U must be equal to or *less* than the tabled value, or, the obtained U' must be equal to or *greater* than

its corresponding critical value. Employing $\alpha = 0.05$, two-tailed test, we find (Table I_3) that for $n_1 = 5$ and $n_2 = 5$, either $U \leq 2$ or $U' \geq 23$ is required to reject H_0. Since our obtained U of $6 > 2$, we may not reject the null hypothesis.

Formula (18.1) may be employed as a check on the calculation of U and U':

$$U = n_1 n_2 - U'. \tag{18.1}$$

The counting technique for arriving at U can become tedious, particularly with large n's, and frequently leads to error. An alternative procedure, which provides identical results, is to assign ranks to the combined groups, as we did before, and then employ either of the following formulas to arrive at U.

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1, \tag{18.2}$$

or

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2, \tag{18.3}$$

where

R_1 = the sum of ranks assigned to the group with a sample size of n_1,

R_2 = the sum of ranks assigned to the group with a sample size of n_2.

Let us suppose that we conducted a study to determine the effects of a drug on the reaction time to a visual stimulus. Since reaction time (and related measures such as latency, time to traverse a runway, etc.) are commonly skewed to the right because of the restriction on the left of the distribution (i.e., no score can be less than zero) and no restrictions on the right (i.e., the score can take on *any* value greater than zero), the Mann-Whitney U-test is selected in preference to the Student t-ratio. The results of the hypothetical study and the computational procedures are shown in Table 18.2.

To check the above calculations, we should first obtain the value of U', employing formula (18.3):

$$U' = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - 39 = 45.$$

We employ formula (18.1) as a check for our calculations:

$$\begin{aligned} U &= n_1 n_2 - U' \\ &= 56 - 45 \\ &= 11. \end{aligned}$$

Table 18.2
The calculation of the Mann-Whitney U employing formula (18.2) (hypothetical data)

Experimental		Control		
Time, milliseconds	Rank	Time, milliseconds	Rank	
140	4	130	1	$U = n_1n_2 + \frac{n_1(n_1+1)}{2} - 81$
147	6	135	2	$= 56 + \frac{8(9)}{2} - 81$
153	8	138	3	$= 56 + 36 - 81 = 11.$
160	10	144	5	
165	11	148	7	
170	13	155	9	
171	14	168	12	
193	15			
$R_1 = 81$, $n_1 = 8$		$R_2 = 39$, $n_2 = 7$		

Employing $\alpha = 0.01$, two-tailed test, for $n_1 = 8$ and $n_2 = 7$, we find (Table I_1) that a $U \leq 6$ is required to reject H_0. Since the obtained U of 11 is greater than this value, we accept H_0.

Tables I_1 through I_4 have been constructed so that it is not necessary to calculate both U and U'. Indeed, it is not even necessary to identify which of these statistics has been calculated. For any given n_1 and n_2, at a specific α level, the tabled values represent the upper and the lower limits of the critical region. The obtained statistic, whether it is actually U or U', must fall *outside* these limits to be significant. Thus you need not be concerned about labeling which of the statistics you have calculated.

18.2.1 Mann-Whitney *U*-Test with Tied Ranks

A problem that often arises with data is that several scores may be exactly the same. Although the underlying dimension on which we base our measures may be continuous, our measures are, for the most part, quite crude. Even though, theoretically, there should be no ties (if we had sufficiently sensitive measuring instruments), we do, as a matter of fact, obtain ties quite often. Although ties within a group do not constitute a problem (U is unaffected), we do face some difficulty when ties occur between two or more observations which involve both groups. This is a formula available which corrects for the effects of ties. Un-

fortunately, the use of this formula is rather involved and is beyond the scope of this introductory text.* However, the failure to correct for ties results in a test which is more "conservative," i.e., decreases the probability of a type I error (rejecting the null hypothesis when it should not be rejected). Correcting for ties is recommended only when their proportion is high and when the uncorrected U approaches our previously set level of significance.

18.3 NONPARAMETRIC TESTS INVOLVING CORRELATED SAMPLES

In Chapter 14, when discussing the Student t-ratio for correlated samples and the algebraically equivalent Sandler A-statistic, we noted the advantages of employing correlated samples wherever feasible. The same advantages accrue to nonparametric tests involving matched or correlated samples. In this section we shall discuss two such tests for ordinally scaled variables, i.e., the *sign test* and the *Wilcoxon signed rank test.*

18.4 THE SIGN TEST

Let us suppose that we are repeating the leadership experiment with which we introduced the chapter, employing larger samples. On the expectation that intelligence and leadership ability are correlated variables, we set up two groups, an experimental and a control, which are matched on the basis of intelligence. On completion of the leadership training course, independent observers are asked to rate the leadership qualities of each subject on a 50-point scale. The results are listed in Table 18.3.

The rating scales seem to be extremely crude and we are unwilling to affirm that the scores have any precise quantitative properties. The only assumption we feel justified in making is that any difference which exists between two paired scores is a valid indicator of the direction and not the magnitude of the difference.

There are 13 pairs of observations in Table 18.3. Since pair M is tied and there is, consequently, no indication of a difference one way or another, we drop these paired observations. Of the remaining 12 pairs, we would expect, on the basis of the null hypothesis, half of the changes to be in the positive direction and half of the changes to be in the negative direction. In other words, under H_0, the probability of any difference being positive is equal to the probability that it will be negative. Since we are dealing with a two-category population (positive differences and negative differences), H_0 may be expressed in pre-

* See Siegel (1956), pp. 123–125, for corrections when a large number of ties occur.

Table 18.3

Ratings of two groups of matched subjects on qualities of leadership (hypothetical data)

	Leadership score		
Matched pair	Experimental	Control	Sign of difference ($E - C$)
A	47	40	+
B	43	38	+
C	36	42	−
D	38	25	+
E	30	29	+
F	22	26	−
G	25	16	+
H	21	18	+
I	14	8	+
J	12	4	+
K	5	7	−
L	9	3	+
M	5	5	(0)

cisely the same fashion as in the binomial test when $P = Q = \frac{1}{2}$. That is, in the present problem, H_0: $P = Q = \frac{1}{2}$. Indeed, the sign test is merely a variation of the binomial test introduced in Section 17.2.

In the present example, out of 12 comparisons showing a difference ($N = 12$), 9 are positive, and 3 are negative. Since $P = Q = \frac{1}{2}$, we refer to Table M, under $x = 3$, $N = 12$, and find that the one-tailed probability is 0.073. The two-tailed probability is therefore 0.146. Employing $\alpha = 0.05$ (two-tailed test) we accept H_0 since $p > 0.05$.

The assumptions underlying the use of the sign test are that the pairs of measurements must be independent of each other and that these measurements must represent, at least, ordinal scaling.

One of the disadvantages of the sign test is that it completely eliminates any quantitative information that may be inherent in the data, (for example, $-8 = -7 = -6$, etc.). The sign test treats all plus differences as if they were the same and all minus differences as if they were the same.

If this is the only assumption warranted by the scale of measurement employed, we have little choice but to employ the sign test. If, on the other hand, the data *do* permit us to make such quantitative statements as "a difference of $8 > 7 > 6 > \cdots$," we lose power when we employ the sign test.

Table 18.4

Ratings of two groups of matched subjects on qualities of leadership (hypothetical data)

Matched pair	Leadership score: Experimental	Control	Difference	Rank of difference	Ranks with smaller sum
A	47	40	+7	9	
B	43	38	+5	5	
C	36	42	−6	−7	−7
D	38	25	+13	12	
E	30	29	+1	1	
F	22	26	−4	−4	−4
G	25	16	+9	11	
H	21	18	+3	3	
I	14	8	+6	7	
J	12	4	+8	10	
K	5	7	−2	−2	−2
L	9	3	+6	7	
M	5	5	(0)	—	
					$T = -13$

18.5 WILCOXON MATCHED-PAIRS SIGNED-RANK TEST

We have seen that the sign test simply utilizes information concerning the direction of the differences between pairs. If the *magnitude* as well as the *direction* of these differences may be considered, a more powerful test may be employed. The *Wilcoxon matched-pairs signed-rank test* achieves greater power by utilizing the quantitative information inherent in the ranking of the differences.

For heuristic purposes, let us return to the data in Table 18.3 and make a different assumption about the scale of measurement employed. Suppose that the rating scale is not as crude as we had imagined, i.e., not only do the measurements achieve ordinal scaling but also the differences between measures achieve ordinality. Table 18.4 reproduces these data, with an additional entry indicating the magnitude of the differences.

Note that the difference column represents differences in scores rather than in ranks. The following column represents the ranking of these differences from smallest to largest without regard to the algebraic sign. Now, if the null hypothesis were correct, we would expect the sum of the positive and that of the negative ranks to more or less balance each other. The more the sum of the ranks are preponderantly positive or negative, the more likely we are to reject the null hypothesis.

The statistic T is the sum of the ranks with the smaller sum. In the above problem, T is equal to -13. Table J presents the critical values of T for sample sizes up to 50 pairs. All entries are for the absolute value of T. In the present example, we find that a T of 13 or less is required for significance at the 0.05 level (two-tailed test) when $n = 12$. Note that we dropped the M-pair from our calculations since, as with the sign test, a zero difference in scores cannot be considered as either a negative or a positive change. Since our obtained T was 13, we may reject the null hypothesis.

You will recall that the sign test applied to these same data did not lead to the rejection of the null hypothesis. The reason should be apparent, i.e., we were not taking advantage of all the information inherent in our data when we employed the sign test.

18.5.1 Assumptions Underlying Wilcoxon's Matched-pairs Signed-Rank Test

An assumption involved in the use of the Wilcoxon signed-rank test is that the scale of measurement is at least ordinal in nature. In other words, the assumption is that the scores permit the ordering of the data into relationships of greater than and less than. However, the signed-rank test makes one additional assumption which may rule it out of some potential applications, namely, that the differences in scores also constitute an ordinal scale. It is not always clear whether or not this assumption is valid for a given set of data. Take, for example, a personality scale purported to measure "manifest anxiety" in a testing situation. Can we validly claim that a difference between matched pairs of, say, 5 points on one part of the scale is greater than a difference of 4 points on another part of the scale? If we cannot validly make this assumption, we must employ another form of statistical analysis, even if it requires that we move to a less sensitive test of significance. Once again, our basic conservatism as scientists makes us more willing to risk a type II rather than a type I error.

CHAPTER SUMMARY

Let us briefly review what we have learned in this chapter.

We have pointed out that the behavioral scientist does not first collect data and then "shop around" for a statistical test to determine the significance of differences between experimental conditions. *The researcher must specify in advance of the experiment* his null hypothesis, alternative hypothesis, test of significance, and the probability value which he will accept as the basis for rejecting the null hypothesis.

We demonstrated the use of the Mann-Whitney U-test as an alternative to the Student t-ratio when the measurements fail to achieve interval scaling or when the researcher wishes to avoid the assumptions of the parametric counter-

part. It is one of the most powerful of the nonparametric tests, since it utilizes most of the quantitative information inherent in the data.

We have seen that by taking into account correlations between subjects on a variable correlated with the criterion measure, we can increase the sensitivity of our statistical test.

The sign test accomplishes this objective by employing before-after measures on the same individuals.

We have also seen that the sign test, although taking advantage of the *direction* of differences involved in ordinal measurement, fails to make use of information concerning *magnitudes* of difference.

The Wilcoxon matched-pairs signed-rank test takes advantage of both *direction* and *magnitude* implicit in ordinal measurement with correlated samples. When the assumptions underlying the test are met, the Wilcoxon paired replicates technique is an extremely sensitive basis for obtaining probability values.

Terms to Remember:

Mann-Whitney U-test
Sign test
Wilcoxon matched-pairs signed-rank test

EXERCISES

1. For the data presented below, determine whether there is a significant difference in the number of stolen bases obtained by two leagues, employing
   a) the sign test, b) the Wilcoxon matched-pairs test,
   c) the Mann-Whitney U-test.
   Which is the best statistical test for these data? Why?

Number of stolen bases		
Team standing	League 1	League 2
1	91	81
2	46	51
3	108	63
4	99	51
5	110	46
6	105	45
7	191	66
8	57	64
9	34	90
10	81	28

2. In a study to determine the effect of a drug on aggressiveness, group A received a drug and group B received a placebo. A test of aggressiveness was applied following the drug administration. The scores obtained were as follows (the higher the score, the greater the aggressiveness):

Group A	10	8	12	16	5	9	7	11	6
Group B	12	15	20	18	13	14	9	16	

Set this study up in formal statistical terms and state the conclusion which is warranted by the statistical evidence.

3. The personnel director at a large insurance office claims that insurance agents who are trained in personal-social relations make more favorable impressions on prospective clients. To test this hypothesis, 22 individuals are randomly selected from those most recently hired and half are assigned to the personal-social relations course. The remaining 11 individuals constitute the control group. Following the training period, all 22 individuals are observed in a simulated interview with a client, and they are rated on a ten-point scale (0–9) for their ease in establishing relationships. The higher the score, the better the rating. Set up and test H_0 employing the appropriate test statistic. Use $\alpha = 0.01$.

Experimentals:	8	7	9	4	7	9	3	7	8	9	3
Controls:	5	6	2	6	0	2	6	5	1	0	5

4. Assume that the subjects in Problem 3 were matched on a variable known to be correlated with the criterion variable. Employ the appropriate test statistic to test H_0: $\alpha = 0.01$.

5. Fifteen husbands and their wives were administered an opinion scale to assess their attitudes about a particular political issue. The results were as follows (the higher the score, the more favorable the attitude):

Husband	Wife	Husband	Wife
37	33	32	46
46	44	35	32
59	48	39	29
17	30	37	45
41	56	36	29
36	30	45	48
29	35	40	35
38	38		

What do you conclude?

Review of Section II. Inferential Statistics

B. NONPARAMETRIC TESTS OF SIGNIFICANCE

In the two preceding chapters you have seen several ways of handling data for which parametric tests of significance were not appropriate. Below are some data, based on a hypothetical experiment. Formulate the null hypothesis, the alternative hypothesis, and conduct the statistical analysis appropriate to the assumptions enumerated in each problem.

Experimental	37	35	33	28	26	20	16	14	12	10	7	5
Control	32	24	29	31	15	18	23	5	9	2	4	1

1. Subjects are randomly assigned to both groups. The scale of measurement is ordinal, and the population is *not* normally distributed.

2. Subjects are matched on a variable known to be correlated with the criterion variable. However, differences in scores *cannot* be assumed to represent *magnitudes* of differences but only direction.

3. Subjects are matched. Scores are based on an ordered scale. Differences in scores may be assumed to be ordinal. However, the population of scores is *not* normally distributed.

4. Determine how effective the matching techniques were.

References

AUBLE, D. (1953), Extended tables for the Mann-Whitney statistic. *Bulletin of the Institute of Educational Research at Indiana University*, **1,** No. 2.

BINGHAM, W. V. (1937), *Aptitudes and Aptitude Testing.* New York: Harper and Bros.

DUNLAP, J. W., and A. K. KURTZ (1932), *Handbook of Statistical Nomographs, Tables, and Formulas.* New York: World Book Company.

EDWARDS, A. L. (1950), *Experimental Design in Psychological Research,* 3rd ed. New York: Rinehart.

EDWARDS, A. L. (1969), *Statistical Analysis,* 3rd ed., New York: Holt, Rinehart and Winston, Inc.

FISHER, R. A. (1935), *The Design of Experiments.* Edinburgh: Oliver and Boyd, Ltd.

FISHER, R. A. (1950), *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd, Ltd.

FISHER, R. A., and F. YATES (1948), *Statistical Tables for Biological, Agricultural, and Medical Research.* Edinburgh: Oliver and Boyd, Ltd.

HABER, AUDREY, and H. I. KALISH (1963), Prediction of Discrimination from Generalization after Variations in Schedule of Reinforcement. *Science,* **142,** 3590, 412–413.

HUFF, D. (1954), *How to Lie with Statistics.* New York: W. W. Norton and Co., Inc.

KIRK, R. E. (1968), *Experimental Design: Procedures for the Behavioral Sciences.* California: Brooks/Cole.

MANN, H. B., and D. R. WHITNEY (1947), On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.,* **18,** 52–54.

McNEMAR, Q. (1962), *Psychological Statistics.* New York: John Wiley and Sons, Inc.

MERRINGTON, M. and C. M. THOMPSON. Tables of Percentage Points of the Inverted Beta Distribution. *Biometrika,* **33,** 73.

OLDS, E. G. (1949), The 5 percent significance levels of sums of squares of rank differences and a correction. *Ann. Math. Statist.,* **20,** 117–118.

PEARSON, E. S. and H. O. HARTLEY (1958), *Biometrika Tables for Statisticians*, Vol. 1, 2nd ed. New York: Cambridge.

RAND CORPORATION (1955), *A Million Random Digits*. Glencoe, Ill.: The Free Press of Glencoe.

RUNYON, R. P. (1968), Note on use of the A statistic as a substitute for t in the one-sample case. *Psychological Reports*, **22**, 361–362.

RUNYON, R. P., HABER, A., and BADIA, P. (1970), *Readings in Statistics*. Reading, Mass.: Addison-Wesley Publishing Co., Inc.

RUNYON, R. P., and M. KOSACOFF (1965), *Olfactory Stimuli as Reinforcers of Bar Pressing Behavior*. Paper presented at Eastern Psych. Assn. meeting.

RUNYON, R. P., and W. J. TURNER (1964), *A Study of the Effects of Drugs on the Social Behavior of White Rats*. New York: Long Island University.

SANDLER, J. (1955), A test of the significance of the difference between the means of correlated measures, based on a simplification of Student's t. *Brit. J. Psychol.*, **46**, 225–226.

SIEGEL, S. (1956), *Non-Parametric Statistics*. New York: McGraw-Hill.

SNEDECOR, G. W., and W. G. COCHRAN (1956), *Statistical Methods*, 6th ed. Ames, Iowa: Iowa State University Press.

TUKEY, J. W. (1953), The Problem of Multiple Comparisons. Ditto, Princeton University, 396 pp.

WALKER, HELEN, and J. LEV (1953), *Statistical Inference*. New York: Henry Holt & Co., Inc.

WILCOXON, F., S. KATTE, and R. A. WILCOX (1963), *Critical Values and Probability Levels for the Wilcoxon Rank Sum Test and the Wilcoxon Signed Rank Test*. New York: American Cyanamid Co.

WILCOXON, F., and R. A. WILCOX (1964), *Some Rapid Approximate Statistical Procedures*. New York: Lederle Laboratories.

Appendixes

	Page

Appendix I
Review of Basic Mathematics

ARITHMETIC OPERATIONS

You already know that addition is indicated by the sign "+," subtraction by the sign "—," multiplication in one of three ways, 2×4, $2(4)$, or $2 \cdot 4$, and division by a slash, "/," a bar, "—," or the symbol "÷." However, it is not unusual to forget the rules concerning addition, subtraction, multiplication, and division, particularly when these operations occur in a single problem.

Addition and Subtraction

When a series of numbers are added together, the order of adding the numbers has no influence on the sum. Thus, we may add $2 + 5 + 3$ in any of the following ways:

$$2+5+3, \qquad 5+2+3, \qquad 2+3+5,$$
$$5+3+2, \qquad 3+2+5, \qquad 3+5+2.$$

When a series of numbers containing both positive and negative signs are added, the order of adding the numbers has no influence on the sum. However, it is often desirable to group together the numbers preceded by positive signs, group together the numbers preceded by negative signs, add each group together separately, and subtract the latter sum from the former. Thus,

$$-2+3+5-4+2+1-8$$

may best be added by grouping in the following ways:

$$\begin{array}{r} +3 \\ +5 \\ +2 \\ +1 \\ \hline +11 \end{array} \qquad \begin{array}{r} -2 \\ -4 \\ -8 \\ \\ \hline -14 \end{array} = -3.$$

Incidentally, to subtract a larger numerical value from a smaller numerical value, as in the above example $(11 - 14)$, we ignore the signs, subtract the smaller number from the larger, and affix a negative sign to the sum. Thus, $-14 + 11 = -3$.

Multiplication

The order in which numbers are multiplied has no effect on the product. In other words,

$$\begin{aligned} 2 \times 3 \times 4 &= 2 \times 4 \times 3 = 3 \times 2 \times 4 \\ &= 3 \times 4 \times 2 = 4 \times 2 \times 3 = 4 \times 3 \times 2 = 24. \end{aligned}$$

When addition, subtraction, and multiplication occur in the same expression, we must develop certain procedures governing which operations are to be performed first.

In the following expression,

$$2 \times 4 + 7 \times 3 - 5$$

multiplication is performed first. Thus, the above expression is equal to

$$\begin{aligned} 1)\quad 2 \times 4 &= 8, & \qquad 2)\quad 8 + 21 - 5 = 24. \\ 7 \times 3 &= 21, \\ -5 &= -5, \end{aligned}$$

We may *not* add first and then multiply. Thus, $2 \times 4 + 7$ is *not* equal to $2(4 + 7)$ or 22.

If a problem involves finding the product of one term multiplied by a second expression which includes two or more terms which are either added or subtracted, we may multiply first and then add, or add first and then multiply. Thus the solution to the following problem becomes

$$\begin{aligned} 8(6 - 4) &= 8 \times 6 - 8 \times 4 \\ &= 48 - 32 \\ &= 16, \end{aligned}$$

or

$$\begin{aligned} 8(6 - 4) &= 8(2) \\ &= 16. \end{aligned}$$

In most cases, however, it is more convenient to reduce the expression within the parentheses first. Thus, generally speaking, the second solution appearing above will be more frequently employed.

Finally, if numbers having like signs are multiplied, the product is always positive; e.g., $(+2) \times (+4) = +8$ and $(-2) \times (-4) = +8$. If numbers bearing unlike signs are multiplied, the product is always negative; e.g.,

$$(+2) \times (-4) = -8$$

and

$$(-2) \times (+4) = -8.$$

The same rule applies also to division: when we obtain the quotient of two numbers of like signs, it is always positive; when the numbers differ in sign, the quotient is always negative.

Multiplication as successive addition. Many students tend to forget that multiplication is a special form of successive addition. Thus

$$15 + 15 + 15 + 15 + 15 = 5(15)$$

and

$$(15 + 15 + 15 + 15 + 15) + (16 + 16 + 16 + 16) = 5(15) + 4(16).$$

This formulation is useful in understanding the advantages of "grouping" scores into what is called a frequency distribution. In obtaining the sum of an array of scores, some of which occur a number of times, it is desirable to multiply each score by the frequency with which it occurs, and then add the products. Thus, if we were to obtain the following distribution of scores,

12, 13, 13, 13, 14, 14, 14, 14, 15, 15, 15, 15,
15, 15, 15, 16, 16, 16, 17, 17, 17, 17, 18,

and wanted the sum of these scores, it would be advantageous to form the following frequency distribution:

X	f	fX
12	1	12
13	3	39
14	4	56
15	7	105
16	3	48
17	4	68
18	1	18
	$N = 23$	$\sum fX = 346$

ALGEBRAIC OPERATIONS

Transposing

To transpose a term from one side of an equation to another, you merely have to change the sign of the transposed term. All the following are equivalent statements:

$$\begin{aligned} a + b &= c, \\ a &= c - b, \\ b &= c - a, \\ 0 &= c - a - b, \\ 0 &= c - (a + b). \end{aligned}$$

Solving Equations Involving Fractions

Much of the difficulty encountered in solving equations which involve fractions can be avoided by remembering one important mathematical principle:

Equals multiplied by equals are equal.

Let us look at a few sample problems.

1. Solve the following equation for x;

$$b = a/x.$$

In solving for x, we want to express the value of x in terms of a and b. In other words, we want our final equation to read, $x =$

Note that we may multiply both sides of the equation by x/b and obtain the following:

$$\cancel{b} \cdot \frac{x}{\cancel{b}} = \frac{a}{\cancel{x}} \cdot \frac{\cancel{x}}{b}.$$

This reduces to

$$x = \frac{a}{b}.$$

2. Solve the above equation for a.

Similarly, if we wanted to solve the equation in terms of a, we could multiply both sides of the equation by x. Thus

$$b \cdot x = \frac{a}{\cancel{x}} \cdot \cancel{x}$$

becomes $bx = a$, or $a = bx$.

In each of the above solutions, you will note that the net effect of multiplying by a constant has been to rearrange the terms in the numerator and the denominator of the equations. In fact, we may state two general rules which

will permit us to solve the above problems without having to employ multiplication by equals (although multiplication by equals is implicit in the arithmetic operations):

a) A term which is in the denominator on one side of the equation may be moved to the other side of the equation by multiplying it by the numerator on that side. Thus

$$\frac{x}{a} = b \qquad \text{becomes} \qquad x = ab.$$

b) A term in the numerator on one side of an equation may be moved to the other side of the equation by dividing it by the numerator on that side. Thus

$$ab = x \qquad \text{may become} \qquad a = \frac{x}{b} \quad \text{or} \quad b = \frac{x}{a}\cdot$$

Thus we have seen that all of the following are equivalent statements:

$$b = \frac{a}{x}, \qquad a = bx, \qquad x = \frac{a}{b}\cdot$$

Similarly,

$$\frac{\sum X}{N} = \overline{X}, \qquad \sum X = N\overline{X}, \qquad \frac{\sum X}{\overline{X}} = N.$$

Dividing by a sum or a difference. It is true that

$$\frac{x+y}{z} = \frac{x}{z} + \frac{y}{z} \qquad \text{and} \qquad \frac{x-y}{z} = \frac{x}{z} - \frac{y}{z}\cdot$$

We cannot, however, simplify the following expressions as easily:

$$\frac{x}{y+z} \qquad \text{or} \qquad \frac{x}{y-z}\cdot$$

Thus,

$$\frac{x}{y+z} \neq \frac{x}{y} + \frac{x}{z},$$

in which $\neq$ means "not equal to."

REDUCING FRACTIONS TO SIMPLEST EXPRESSIONS

A corollary to the rule that equals multiplied by equals are equal is:

Unequals multiplied by equals remain proportional.

Thus, if we were to multiply $\frac{1}{4}$ by $\frac{8}{8}$, the product, $\frac{8}{32}$, is in the same proportion

as $\frac{1}{4}$. This corollary is useful in reducing the complex fractions to their simplest expression. Let us look at an example.

Example: Reduce
$$\frac{\frac{a}{b}}{\frac{c}{d}} \quad \text{or} \quad \frac{a}{b} \div \frac{c}{d}$$
to its simplest expression.

Note that if we multiply both the numerator and the denominator by
$$\frac{\frac{bd}{1}}{\frac{bd}{1}}$$
we obtain
$$\frac{\frac{a}{\cancel{b}} \cdot \frac{\cancel{b}d}{1}}{\frac{c}{\cancel{d}} \cdot \frac{b\cancel{d}}{1}},$$
which becomes ad/bc.

However, we could obtain the same result if we were to invert the divisor and multiply. Thus
$$\frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a}{b} \cdot \frac{d}{c} = \frac{ad}{bc}.$$

We may now formulate a general rule for dividing one fraction into another fraction. In dividing fractions, we invert the divisor and multiply. Thus
$$\frac{\frac{x}{y}}{\frac{a^2}{b}} \quad \text{becomes} \quad \frac{x}{y} \cdot \frac{b}{a^2} \quad \text{which equals} \quad \frac{bx}{a^2y}.$$

To illustrate: If $a = 5$, $b = 2$, $x = 3$, and $y = 4$, the above expressions become
$$\frac{\frac{3}{4}}{\frac{5^2}{2}} = \frac{3}{4} \cdot \frac{2}{5^2} = \frac{2 \cdot 3}{4 \times 5^2} = \frac{6}{100}.$$

A general practice you should follow when substituting numerical values into fractional expressions is to reduce the expression to its simplest form *prior* to substitution.

Multiplication and Division of Terms Having Exponents

An exponent indicates how many times a number is to be multiplied by itself. For example, X^5 means that X is to be multiplied by itself 5 times, or

$$X^5 = X \cdot X \cdot X \cdot X \cdot X.$$

If $X = 3$,

$$X^5 = 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 243 \qquad \text{and} \qquad \left(\frac{1}{X}\right)^5 = \frac{1^5}{X^5} = \frac{1 \cdot 1 \cdot 1 \cdot 1 \cdot 1}{3 \cdot 3 \cdot 3 \cdot 3 \cdot 3} = \frac{1}{243}\cdot$$

To multiply X raised to the ath power (X^a) times X raised to the bth power, you simply add the exponents and raise X to the $(a + b)$th power. The reason for the addition of exponents may be seen from the following illustration.

If $a = 3$ and $b = 5$, then

$$X^a \cdot X^b = X^3 X^5 = (X \cdot X \cdot X)(X \cdot X \cdot X \cdot X \cdot X),$$

which equals X^8.

Now, if $X = 5$, $a = 3$, and $b = 5$, then

$$X^a \cdot X^b = X^{a+b} = X^{3+5} = X^8 = 5^8 = 390{,}625.$$

If $X = \frac{1}{6}$, $a = 2$, and $b = 3$,

$$X^a \cdot X^b = X^{a+b} = \left(\frac{1}{6}\right)^{2+3} = \left(\frac{1}{6}\right)^5 = \frac{1^5}{6^5} = \frac{1}{7776}\cdot$$

To divide X raised to the ath power by X raised to the bth power, you simply subtract the exponent in the denominator from the exponent in the numerator.* The reason for the subtraction is made clear in the following illustration.

---

* This leads to an interesting exception to the rule that an exponent indicates the number of times a number is multiplied by itself; that is,

$$\frac{X^n}{X^n} = X^{n-n} = X^0;$$

however,

$$\frac{X^n}{X^n} = 1; \qquad \text{therefore} \qquad X^0 = 1.$$

Any number raised to the zero power is equal to 1.

If $X = 3$, $a = 5$, and $b = 2$, then

$$\frac{X^a}{X^b} = \frac{X^5}{X^2} = \frac{X \cdot X \cdot X \cdot X \cdot X}{X \cdot X} = X^3 = 3^3 = 27.$$

If $X = \frac{5}{6}$, $a = 4$, $b = 2$, then

$$\frac{X^a}{X^b} = X^{a-b} = X^{4-2} = X^2.$$

Substituting $\frac{5}{6}$ for X, we have

$$X^2 = \left(\frac{5}{6}\right)^2 = \frac{5^2}{6^2} = \frac{25}{36}.$$

EXTRACTING SQUARE ROOTS

The square root of a number is the value which, when multiplied by itself, equals that number. Appendix III contains a table of square roots.

The usual difficulty encountered in calculating square roots is the decision as to how many digits precede the decimal, for example $\sqrt{25{,}000{,}000} = 5000$, not 500 or 50,000; i.e., there are four digits before the decimal. In order to calculate the number of digits preceding the decimal, simply count the number of pairs to the left of the decimal:

$$\text{number of pairs} = \text{number of digits.}$$

However, if there is an odd number of digits, then the number of digits preceding the decimal equals the number of pairs +1. The following examples illustrate this point:

$$\text{a)}\quad \overset{50.0}{\sqrt{2500.00}}, \qquad \overset{5.0}{\sqrt{25.00}};$$

$$\text{b)}\quad \overset{15.8}{\sqrt{250.00}}, \qquad \overset{1.58}{\sqrt{2.5000}}.$$

Appendix II
Glossary of Symbols

Listed below are definitions of the symbols which appear in the text followed by the page number showing the first reference to the symbol.

English letters and Greek letters are listed separately in their approximate alphabetical order. Mathematical operators are also listed separately.

Symbol	Definition
	MATHEMATICAL OPERATORS
$\neq$	Not equal to (14)
$a < b$	a is less than b (14)
$a > b$	a is greater than b (14)
$\leq$	Less than or equal to (138)
$\geq$	Greater than or equal to (138)
$\sqrt{\ }$	Square root (10)
X^a	X raised to the ath power (10)
$N!$	Factorial: multiply N by all integers less than it but greater than zero: $(N)\,(N-1)\,(N-2)\cdots(2)\,(1)$ (244)
$\lvert X\rvert$	Absolute value of X (71)
$\sum$	Sum all quantities or scores that follow (10)
$\sum_{i=1}^{N} X_i$	Sum all quantities X_1 through X_N: $X_1 + X_2 + \cdots + X_N$ (11)

Symbol	Definition
	GREEK LETTERS
α	Probability of a type I error, probability of rejecting H_0 when it is true (165)
β	Probability of a type II error, probability of accepting H_0 when it is false (168)
χ^2	Chi square (247)

μ	Population mean (57)
μ_0	Value of the population mean under H_0 (176)
μ_1	Value of the population mean under H_1 (232)
$\mu_{\overline{X}}$	Mean of the distribution of sample means (175)
$\mu_{\overline{X}_1-\overline{X}_2}$	Mean of the distribution of the difference between pairs of sample means (195)
μ_D	Mean of the difference between paired scores (207)
σ^2	Population variance (73)
σ	Population standard deviation (73)
$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{N}$	Variance of the sampling distribution of the mean (175)
$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$	True standard error of the mean given random samples of a fixed N (175)
$\sigma_{\overline{X}_1-\overline{X}_2}$	True standard error of the difference between means (195)

Symbol	Definition
	ENGLISH LETTERS
$A = \frac{\sum D^2}{(\sum D)^2}$	Statistic employed to test hypotheses for correlated samples (210)
a	Constant term in a regression equation (112)
b_y	Slope of a line relating values of Y to values of X (112)
c	Number of columns in a contingency table (250)
cum f	Cumulative frequency (28)
cum f_{ll}	Cumulative frequency at the lower real limit of the interval containing X (50)
cum %	Cumulative percent (28)
D	(1) Rank on X-variable − rank on Y-variable (r_{rho} formula) (103) (2) Score on X-variable − score on Y-variable $(X - Y)$ (208)
$\overline{D}$	Mean of the differences between the paired scores (207)
d	Deviation of a difference score (D) from $\overline{D}$ (208)
df	Degrees of freedom: number of values free to vary after certain restrictions have been placed on the data (180)
F	A ratio of two variances (199)
f	Frequency (25)

f_i	Number of cases within the interval containing X (50)
f_e	Expected number in a given category (247)
f_o	Observed number in a given category (247)
fX	A score multiplied by its corresponding frequency (58)
H_0	The null hypothesis; hypothesis actually tested (166)
H_1	The alternative hypothesis; hypothesis entertained if H_0 is rejected (166)
i	Width of the class interval (26)
k	Number of groups or categories (219)
k^2	Coefficient of nondetermination (123)
M.D.	Mean deviation (71)
N	Total number of scores or quantities (10)
n	(1) Number of pairs (97) (2) Number in either sample (195)
p	(1) Proportion (20) (2) Probability (137)
$p(A)$	Probability of event A (137)
$p(B/A)$	Probability of B given than A has occurred (141)
P	(1) Probability of the occurrence of an event (140) (2) Proportion of cases in one class in a two-category population (243)
Q	(1) Probability of the nonoccurrence of an event (140) (2) Proportion of cases in the other class of a two-category population (243)
Q_1	First quartile, 25th percentile (71)
Q_3	Third quartile, 75th percentile (71)
r	(1) Pearson product-moment correlation coefficient (94) (2) Number of rows in a contingency table (250)
r^2	Coefficient of determination (122)
r_{rho}	Spearman rank-order correlation coefficient (94)
R_1	Sum of ranks assigned to the group with a sample size of n_1 (Mann-Whitney U-formula) (259)
R_2	Sum of ranks assigned to the group with a sample size of n_2 (Mann-Whitney U-formula) (259)
$s^2 = \dfrac{\sum x^2}{N}$	Variance of a sample (73)

$s = \sqrt{\frac{\sum x^2}{N}}$ Standard deviation of a sample (73)

$\hat{s}^2 = \frac{\sum x^2}{N - 1}$ Unbiased estimate of the population variance (178)

$\hat{s} = \sqrt{\frac{\sum x^2}{N - 1}}$ Sample standard deviation based on unbiased variance estimate (178)

$s_{\overline{X}}^2$ Estimated variance of the sampling distribution of the mean (178)

$s_{\overline{X}} = \frac{\hat{s}}{\sqrt{N}} = \frac{s}{\sqrt{N - 1}}$ Estimated standard error of the mean (178)

$s_{\overline{X}_1 - \overline{X}_2}$ Estimated standard error of the difference between means (195)

$\hat{s}_D$ Standard deviation of the difference scores (208)

$s_{\bar{D}}$ Estimated standard error of the difference between means, direct-difference method (208)

$\hat{s}_B^2$ Between-group variance estimate (219)

$\hat{s}_W^2$ Within-group variance estimate (219)

$s_{\text{est } y}$ Standard error of estimate when prediction are made from X to Y (119)

$s_{\text{est } x}$ Standard error of estimate when predictions are made from Y to X (119)

T Sum of the ranks with the least frequent sign (264)

t Statistic employed to test hypotheses when σ is unknown (180)

U, U' Statistics in the Mann-Whitney test (258)

X, Y Variables; quantities or scores of variables (10)

X_i, Y_i Specific quantities indicated by the subscript i (10)

$\overline{X}, \overline{Y}$ Arithmetic means (57)

$\overline{X}_i$ Mean of the ith group (218)

$\overline{X}_{\text{tot}} = \frac{\sum X_{\text{tot}}}{N}$ Overall mean (218)

$x = X - \overline{X}$ Deviation of a score from its mean (74)

x Number of objects in one category or the number of successes (244)

$\sum X^2$ Sum of the squares of the raw scores (75)

$(\sum X)^2$ Sum of the raw scores, the quantity squared (75)

$\sum x^2 = \sum(X - \overline{X})^2$ Sum squares, sum of the squared deviations from the mean (75)

$\sum x_{\text{tot}}^2$	Total sum squares, sum of the squared deviations of each score (X) from the overall mean ($\overline{X}_{\text{tot}}$) (217)
$\sum x_W^2$	Within-group sum squares, sum of the squared deviations of each score (X) from the mean of its own group ($\overline{X}_i$) (217)
$\sum x_B^2$	Between-group sum squares, sum of the squared deviations of each group mean ($\overline{X}_i$) from the overall mean ($\overline{X}_{\text{tot}}$), multiplied by the n in each group (217)
X_{ll}	Score at lower real limit of interval containing X (50)
X', Y'	Scores predicted by regression equations (113)
z	(1) Deviation of a specific score from the mean expressed in standard deviation units (82) (2) Statistic employed to test hypotheses when σ is known (176, 195)
$z_{0.01} = \pm 2.58$	Critical value of z, minimum z required to reject H_0 at the 0.01 level of significance, two-tailed test (177)
$z_{0.05} = \pm 1.96$	Minimum value of z required to reject H_0 at the 0.05 level of significance, two-tailed test (177)
$z_{y'}$	Y' expressed in terms of a z-score (116)

Appendix III
Tables

ACKNOWLEDGMENTS

The authors are grateful to the authors and publishers listed below for permission to adapt from the following tables.

Table B Table III of R. A. FISHER and F. YATES, *Methods for Research Workers*. Edinburgh: Oliver and Boyd, Ltd., 1950.

Table C Table III of R. A. FISHER and F. YATES, *Statistical Tables for Biological, Agricultural, and Medical Research*. Edinburgh: Oliver and Boyd, Ltd., 1948.

Table D G. W. SNEDECOR and WILLIAM G. COCHRAN, *Statistical Methods*, 6th edition, Ames, Iowa: Iowa State University Press © 1967.

Table D_1 M. MERRINGTON and C. M. THOMPSON, "Tables of Percentage Points of the Inverted Beta Distribution." *Biometrika*, **33**, 73, 1943.

Table E J. SANDLER, A test of the significance of the difference between the means of correlated measures, based on a simplification of Student's *t*. *Brit. J. Psychol.*, **46,** 225–226, 1955.

Table F Q. McNEMAR, Table B of *Psychological Statistics*. New York: John Wiley and Sons, Inc., 1962.

Table G E. G. OLDS, The 5 percent significance levels of sums of squares of rank differences and a correction. *Ann. Math. Statist.*, **20,** 117–118, 1949.
E. G. OLDS, Distribution of sums of squares of rank differences for small numbers of individuals. *Ann. Math. Statist.*, **9,** 133–148, 1938.

Table H W. V. BINGHAM, Table XVII of *Aptitudes and Aptitude Testing*. New York: Harper and Bros., 1937.

Table I H. B. MANN and D. R. WHITNEY, On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, **18,** 52–54, 1947.
D. AUBLE, Extended tables for the Mann-Whitney statistic. *Bulletin of the Institute of Educational Research at Indiana University*, **1,** No. 2, 1953.

Table J F. WILCOXON, S. KATTI, and R. A. WILCOX, *Critical Values and Probability Levels for the Wilcoxon Rank Sum Test and the Wilcoxon Signed Rank test*. New York: American Cyanamid Co., 1963.
F. WILCOXON and R. A. WILCOX, *Some Rapid Approximate Statistical Procedures*. New York: Lederle Laboratories, 1964.

Table L S. SIEGEL, *Non-Parametric Statistics*. New York: McGraw-Hill, 1956.

Table M Table IV B of HELEN WALKER and J. LEV, *Statistical Inference*. New York: Henry Holt & Co., Inc., 1953.

Table P E. S. PEARSON and H. O. HARTLEY, *Biometrika Tables for Statisticians*, Vol. 1, 2nd ed. New York: Cambridge, 1958.

Table Q A. L. EDWARDS, *Statistical Analysis*, 3rd ed. New York: Holt, Rinehart and Winston, Inc., 1969.

J. W. Dunlap and A. K. Kurtz, *Handbook of Statistical Nomographs, Tables, and Formulas*. New York: World Book Company, 1932.

Table R RAND Corporation, *A Million Random Digits*. Glencoe, Ill.: Free Press of Glencoe, 1955.

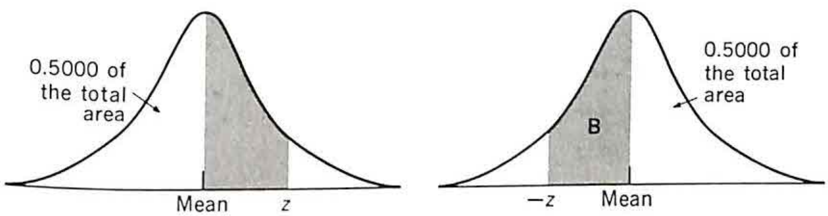
THE USE OF TABLE A

The use of Table A requires that the raw score be transformed into a z-score and that the variable be normally distributed.

The values in Table A represent the proportion of area in the standard normal curve which has a mean of 0, a standard deviation of 1.00, and a total area also equal to 1.00.

Since the normal curve is symmetrical, it is sufficient to indicate only the areas corresponding to positive z-values. Negative z-values will have precisely the same proportions of area as their positive counterparts.

Column B represents the proportion of area between the mean and a given z.



Column C represents the proportion of area beyond a given z.

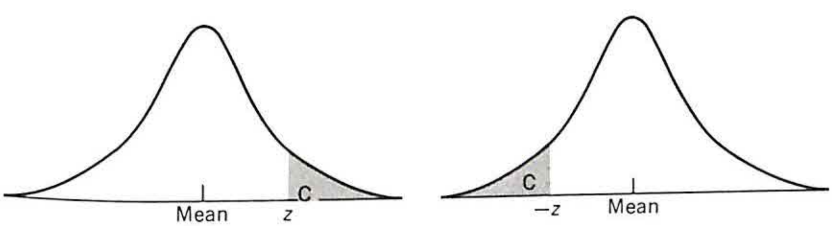


Table A

Proportions of Area under the Normal Curve

(A) z	(B) area between mean and z	(C) area beyond z	(A) z	(B) area between mean and z	(C) area beyond z	(A) z	(B) area between mean and z	(C) area beyond z
0.00	.0000	.5000	0.55	.2088	.2912	1.10	.3643	.1357
0.01	.0040	.4960	0.56	.2123	.2877	1.11	.3665	.1335
0.02	.0080	.4920	0.57	.2157	.2843	1.12	.3686	.1314
0.03	.0120	.4880	0.58	.2190	.2810	1.13	.3708	.1292
0.04	.0160	.4840	0.59	.2224	.2776	1.14	.3729	.1271
0.05	.0199	.4801	0.60	.2257	.2743	1.15	.3749	.1251
0.06	.0239	.4761	0.61	.2291	.2709	1.16	.3770	.1230
0.07	.0279	.4721	0.62	.2324	.2676	1.17	.3790	.1210
0.08	.0319	.4681	0.63	.2357	.2643	1.18	.3810	.1190
0.09	.0359	.4641	0.64	.2389	.2611	1.19	.3830	.1170
0.10	.0398	.4602	0.65	.2422	.2578	1.20	.3849	.1151
0.11	.0438	.4562	0.66	.2454	.2546	1.21	.3869	.1131
0.12	.0478	.4522	0.67	.2486	.2514	1.22	.3888	.1112
0.13	.0517	.4483	0.68	.2517	.2483	1.23	.3907	.1093
0.14	.0557	.4443	0.69	.2549	.2451	1.24	.3925	.1075
0.15	.0596	.4404	0.70	.2580	.2420	1.25	.3944	.1056
0.16	.0636	.4364	0.71	.2611	.2389	1.26	.3962	.1038
0.17	.0675	.4325	0.72	.2642	.2358	1.27	.3980	.1020
0.18	.0714	.4286	0.73	.2673	.2327	1.28	.3997	.1003
0.19	.0753	.4247	0.74	.2704	.2296	1.29	.4015	.0985
0.20	.0793	.4207	0.75	.2734	.2266	1.30	.4032	.0968
0.21	.0832	.4168	0.76	.2764	.2236	1.31	.4049	.0951
0.22	.0871	.4129	0.77	.2794	.2206	1.32	.4066	.0934
0.23	.0910	.4090	0.78	.2823	.2177	1.33	.4082	.0918
0.24	.0948	.4052	0.79	.2852	.2148	1.34	.4099	.0901
0.25	.0987	.4013	0.80	.2881	.2119	1.35	.4115	.0885
0.26	.1026	.3974	0.81	.2910	.2090	1.36	.4131	.0869
0.27	.1064	.3936	0.82	.2939	.2061	1.37	.4147	.0853
0.28	.1103	.3897	0.83	.2967	.2033	1.38	.4162	.0838
0.29	.1141	.3859	0.84	.2995	.2005	1.39	.4177	.0823
0.30	.1179	.3821	0.85	.3023	.1977	1.40	.4192	.0808
0.31	.1217	.3783	0.86	.3051	.1949	1.41	.4207	.0793
0.32	.1255	.3745	0.87	.3078	.1922	1.42	.4222	.0778
0.33	.1293	.3707	0.88	.3106	.1894	1.43	.4236	.0764
0.34	.1331	.3669	0.89	.3133	.1867	1.44	.4251	.0749
0.35	.1368	.3632	0.90	.3159	.1841	1.45	.4265	.0735
0.36	.1406	.3594	0.91	.3186	.1814	1.46	.4279	.0721
0.37	.1443	.3557	0.92	.3212	.1788	1.47	.4292	.0708
0.38	.1480	.3520	0.93	.3238	.1762	1.48	.4306	.0694
0.39	.1517	.3483	0.94	.3264	.1736	1.49	.4319	.0681
0.40	.1554	.3446	0.95	.3289	.1711	1.50	.4332	.0668
0.41	.1591	.3409	0.96	.3315	.1685	1.51	.4345	.0655
0.42	.1628	.3372	0.97	.3340	.1660	1.52	.4357	.0643
0.43	.1664	.3336	0.98	.3365	.1635	1.53	.4370	.0630
0.44	.1700	.3300	0.99	.3389	.1611	1.54	.4382	.0618
0.45	.1736	.3264	1.00	.3413	.1587	1.55	.4394	.0606
0.46	.1772	.3228	1.01	.3438	.1562	1.56	.4406	.0594
0.47	.1808	.3192	1.02	.3461	.1539	1.57	.4418	.0582
0.48	.1844	.3156	1.03	.3485	.1515	1.58	.4429	.0571
0.49	.1879	.3121	1.04	.3508	.1492	1.59	.4441	.0559
0.50	.1915	.3085	1.05	.3531	.1469	1.60	.4452	.0548
0.51	.1950	.3050	1.06	.3554	.1446	1.61	.4463	.0537
0.52	.1985	.3015	1.07	.3577	.1423	1.62	.4474	.0526
0.53	.2019	.2981	1.08	.3599	.1401	1.63	.4484	.0516
0.54	.2054	.2946	1.09	.3621	.1379	1.64	.4495	.0505

Table A. *(continued)*

(A) z	(B) area between mean and z	(C) area beyond z
1.65	.4505	.0495
1.66	.4515	.0485
1.67	.4525	.0475
1.68	.4535	.0465
1.69	.4545	.0455
1.70	.4554	.0446
1.71	.4564	.0436
1.72	.4573	.0427
1.73	.4582	.0418
1.74	.4591	.0409
1.75	.4599	.0401
1.76	.4608	.0392
1.77	.4616	.0384
1.78	.4625	.0375
1.79	.4633	.0367
1.80	.4641	.0359
1.81	.4649	.0351
1.82	.4656	.0344
1.83	.4664	.0336
1.84	.4671	.0329
1.85	.4678	.0322
1.86	.4686	.0314
1.87	.4693	.0307
1.88	.4699	.0301
1.89	.4706	.0294
1.90	.4713	.0287
1.91	.4719	.0281
1.92	.4726	.0274
1.93	.4732	.0268
1.94	.4738	.0262
1.95	.4744	.0256
1.96	.4750	.0250
1.97	.4756	.0244
1.98	.4761	.0239
1.99	.4767	.0233
2.00	.4772	.0228
2.01	.4778	.0222
2.02	.4783	.0217
2.03	.4788	.0212
2.04	.4793	.0207
2.05	.4798	.0202
2.06	.4803	.0197
2.07	.4808	.0192
2.08	.4812	.0188
2.09	.4817	.0183
2.10	.4821	.0179
2.11	.4826	.0174
2.12	.4830	.0170
2.13	.4834	.0166
2.14	.4838	.0162
2.15	.4842	.0158
2.16	.4846	.0154
2.17	.4850	.0150
2.18	.4854	.0146
2.19	.4857	.0143
2.20	.4861	.0139
2.21	.4864	.0136

(A) z	(B) area between mean and z	(C) area beyond z
2.22	.4868	.0132
2.23	.4871	.0129
2.24	.4875	.0125
2.25	.4878	.0122
2.26	.4881	.0119
2.27	.4884	.0116
2.28	.4887	.0113
2.29	.4890	.0110
2.30	.4893	.0107
2.31	.4896	.0104
2.32	.4898	.0102
2.33	.4901	.0099
2.34	.4904	.0096
2.35	.4906	.0094
2.36	.4909	.0091
2.37	.4911	.0089
2.38	.4913	.0087
2.39	.4916	.0084
2.40	.4918	.0082
2.41	.4920	.0080
2.42	.4922	.0078
2.43	.4925	.0075
2.44	.4927	.0073
2.45	.4929	.0071
2.46	.4931	.0069
2.47	.4932	.0068
2.48	.4934	.0066
2.49	.4936	.0064
2.50	.4938	.0062
2.51	.4940	.0060
2.52	.4941	.0059
2.53	.4943	.0057
2.54	.4945	.0055
2.55	.4946	.0054
2.56	.4948	.0052
2.57	.4949	.0051
2.58	.4951	.0049
2.59	.4952	.0048
2.60	.4953	.0047
2.61	.4955	.0045
2.62	.4956	.0044
2.63	.4957	.0043
2.64	.4959	.0041
2.65	.4960	.0040
2.66	.4961	.0039
2.67	.4962	.0038
2.68	.4963	.0037
2.69	.4964	.0036
2.70	.4965	.0035
2.71	.4966	.0034
2.72	.4967	.0033
2.73	.4968	.0032
2.74	.4969	.0031
2.75	.4970	.0030
2.76	.4971	.0029
2.77	.4972	.0028
2.78	.4973	.0027

(A) z	(B) area between mean and z	(C) area beyond z
2.79	.4974	.0026
2.80	.4974	.0026
2.81	.4975	.0025
2.82	.4976	.0024
2.83	.4977	.0023
2.84	.4977	.0023
2.85	.4978	.0022
2.86	.4979	.0021
2.87	.4979	.0021
2.88	.4980	.0020
2.89	.4981	.0019
2.90	.4981	.0019
2.91	.4982	.0018
2.92	.4982	.0018
2.93	.4983	.0017
2.94	.4984	.0016
2.95	.4984	.0016
2.96	.4985	.0015
2.97	.4985	.0015
2.98	.4986	.0014
2.99	.4986	.0014
3.00	.4987	.0013
3.01	.4987	.0013
3.02	.4987	.0013
3.03	.4988	.0012
3.04	.4988	.0012
3.05	.4989	.0011
3.06	.4989	.0011
3.07	.4989	.0011
3.08	.4990	.0010
3.09	.4990	.0010
3.10	.4990	.0010
3.11	.4991	.0009
3.12	.4991	.0009
3.13	.4991	.0009
3.14	.4992	.0008
3.15	.4992	.0008
3.16	.4992	.0008
3.17	.4992	.0008
3.18	.4993	.0007
3.19	.4993	.0007
3.20	.4993	.0007
3.21	.4993	.0007
3.22	.4994	.0006
3.23	.4994	.0006
3.24	.4994	.0006
3.25	.4994	.0006
3.30	.4995	.0005
3.35	.4996	.0004
3.40	.4997	.0003
3.45	.4997	.0003
3.50	.4998	.0002
3.60	.4998	.0002
3.70	.4999	.0001
3.80	.4999	.0001
3.90	.49995	.00005
4.00	.49997	.00003

Table B

Table of χ^2

Degrees of freedom df	P = .99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01
1	.000157	.000628	.00393	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086
6	.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892

Table C

Critical Values of t

For any given df, the table shows the values of t corresponding to various levels of probability. Obtained t is significant at a given level if it is equal to or greater than the value shown in the table.

	Level of significance for one-tailed test					
	.10	.05	.025	.01	.005	.0005
	Level of significance for two-tailed test					
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Table D. Critical Values of *F*

The obtained F is significant at a given level if it is equal to or <u>greater than</u> the value shown in the table.
0.05 (light row) and 0.01 (dark row) points for the distribution of F

Degrees of freedom for lesser mean square	Degrees of freedom for greater mean square																							
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254
	4052	**4999**	**5403**	**5625**	**5764**	**5859**	**5928**	**5981**	**6022**	**6056**	**6082**	**6106**	**6142**	**6169**	**6208**	**6234**	**6258**	**6286**	**6302**	**6323**	**6334**	**6352**	**6361**	**6366**
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50
	98.49	**99.01**	**99.17**	**99.25**	**99.30**	**99.33**	**99.34**	**99.36**	**99.38**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.49**	**99.49**	**99.49**	**99.50**	**99.50**
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53
	34.12	**30.81**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.30**	**26.27**	**26.23**	**26.18**	**26.14**	**26.12**
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.54**	**14.45**	**14.37**	**14.24**	**14.15**	**14.02**	**13.93**	**13.83**	**13.74**	**13.69**	**13.61**	**13.57**	**13.52**	**13.48**	**13.46**
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.27**	**10.15**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.07**	**9.04**	**9.02**
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.39**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.94**	**6.90**	**6.88**
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**	**6.54**	**6.47**	**6.35**	**6.27**	**6.15**	**6.07**	**5.98**	**5.90**	**5.85**	**5.78**	**5.75**	**5.70**	**5.67**	**5.65**
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**	**5.74**	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.11**	**5.06**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.35**	**5.26**	**5.18**	**5.11**	**5.00**	**4.92**	**4.80**	**4.73**	**4.64**	**4.56**	**4.51**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.21**	**5.06**	**4.95**	**4.85**	**4.78**	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.63**	**4.54**	**4.46**	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.80**	**3.74**	**3.70**	**3.66**	**3.62**	**3.60**
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.39**	**4.30**	**4.22**	**4.16**	**4.05**	**3.98**	**3.86**	**3.78**	**3.70**	**3.61**	**3.56**	**3.49**	**3.46**	**3.41**	**3.38**	**3.36**
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**	**4.02**	**3.96**	**3.85**	**3.78**	**3.67**	**3.59**	**3.51**	**3.42**	**3.37**	**3.30**	**3.27**	**3.21**	**3.18**	**3.16**
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**	**3.86**	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.34**	**3.26**	**3.21**	**3.14**	**3.11**	**3.06**	**3.02**	**3.00**
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**	**3.73**	**3.67**	**3.56**	**3.48**	**3.36**	**3.29**	**3.20**	**3.12**	**3.07**	**3.00**	**2.97**	**2.92**	**2.89**	**2.87**

Degrees of freedom for lesser mean square																								
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.45 3.61	2.42 3.55	2.37 3.45	2.33 3.37	2.28 3.25	2.24 3.18	2.20 3.10	2.16 3.01	2.13 2.96	2.09 2.89	2.07 2.86	2.04 2.80	2.02 2.77	2.01 2.75
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.62 3.93	2.55 3.79	2.50 3.68	2.45 3.59	2.41 3.52	2.38 3.45	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.11 2.92	2.08 2.86	2.04 2.79	2.02 2.76	1.99 2.70	1.97 2.67	1.96 2.65
18	4.41 8.28	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.58 3.85	2.51 3.71	2.46 3.60	2.41 3.51	2.37 3.44	2.34 3.37	2.29 3.27	2.25 3.19	2.19 3.07	2.15 3.00	2.11 2.91	2.07 2.83	2.04 2.78	2.00 2.71	1.98 2.68	1.95 2.62	1.93 2.59	1.92 2.57
19	4.38 8.18	3.52 5.93	3.13 5.01	2.90 4.50	2.74 4.17	2.63 3.94	2.55 3.77	2.48 3.63	2.43 3.52	2.38 3.43	2.34 3.36	2.31 3.30	2.26 3.19	2.21 3.12	2.15 3.00	2.11 2.92	2.07 2.84	2.02 2.76	2.00 2.70	1.96 2.63	1.94 2.60	1.91 2.54	1.90 2.51	1.88 2.49
20	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.52 3.71	2.45 3.56	2.40 3.45	2.35 3.37	2.31 3.30	2.28 3.23	2.23 3.13	2.18 3.05	2.12 2.94	2.08 2.86	2.04 2.77	1.99 2.69	1.96 2.63	1.92 2.56	1.90 2.53	1.87 2.47	1.85 2.44	1.84 2.42
21	4.32 8.02	3.47 5.78	3.07 4.87	2.84 4.37	2.68 4.04	2.57 3.81	2.49 3.65	2.42 3.51	2.37 3.40	2.32 3.31	2.28 3.24	2.25 3.17	2.20 3.07	2.15 2.99	2.09 2.88	2.05 2.80	2.00 2.72	1.96 2.63	1.93 2.58	1.80 2.51	1.87 2.47	1.84 2.42	1.82 2.38	1.81 2.36
22	4.30 7.94	3.44 5.72	3.05 4.82	2.82 4.31	2.66 3.99	2.55 3.76	2.47 3.59	2.40 3.45	2.35 3.35	2.30 3.26	2.26 3.18	2.23 3.12	2.18 3.02	2.13 2.94	2.07 2.83	2.03 2.75	1.98 2.67	1.93 2.58	1.91 2.53	1.87 2.46	1.84 2.42	1.81 2.37	1.80 2.33	1.78 2.31
23	4.28 7.88	3.42 5.66	3.03 4.76	2.80 4.26	2.64 3.94	2.53 3.71	2.45 3.54	2.38 3.41	2.32 3.30	2.28 3.21	2.24 3.14	2.20 3.07	2.14 2.97	2.10 2.89	2.04 2.78	2.00 2.70	1.96 2.62	1.91 2.53	1.88 2.48	1.84 2.41	1.82 2.37	1.79 2.32	1.77 2.28	1.76 2.26
24	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.43 3.50	2.36 3.36	2.30 3.25	2.26 3.17	2.22 3.09	2.18 3.03	2.13 2.93	2.09 2.85	2.02 2.74	1.98 2.66	1.94 2.58	1.89 2.49	1.86 2.44	1.82 2.36	1.80 2.33	1.76 2.27	1.74 2.23	1.73 2.21
25	4.24 7.77	3.38 5.57	2.99 4.68	2.76 4.18	2.60 3.86	2.49 3.63	2.41 3.46	2.34 3.32	2.28 3.21	2.24 3.13	2.20 3.05	2.16 2.99	2.11 2.89	2.06 2.81	2.00 2.70	1.96 2.62	1.92 2.54	1.87 2.45	1.84 2.40	1.80 2.32	1.77 2.29	1.74 2.23	1.72 2.19	1.71 2.17
26	4.22 7.72	3.37 5.53	2.89 4.64	2.74 4.14	2.59 3.82	2.47 3.59	2.39 3.42	2.32 3.29	2.27 3.17	2.22 3.09	2.18 3.02	2.15 2.96	2.10 2.86	2.05 2.77	1.99 2.66	1.95 2.58	1.90 2.50	1.85 2.41	1.82 2.36	1.78 2.28	1.76 2.25	1.72 2.19	1.70 2.15	1.69 2.13
27	4.21 7.68	3.35 5.49	2.96 4.60	2.73 4.11	2.57 3.79	2.46 3.56	2.37 3.39	2.30 3.26	2.25 3.14	2.20 3.06	2.16 2.98	2.13 2.93	2.08 2.83	2.03 2.74	1.97 2.63	1.93 2.55	1.88 2.47	1.84 2.38	1.80 2.33	1.76 2.25	1.74 2.21	1.71 2.16	1.68 2.12	1.67 2.10
28	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.76	2.44 3.53	2.36 3.36	2.29 3.23	3.24 3.11	2.19 3.03	2.15 2.95	2.12 2.90	2.06 2.80	2.02 2.71	1.96 2.60	1.91 2.52	1.87 2.44	1.81 2.35	1.78 2.30	1.75 2.22	1.72 2.18	1.69 2.13	1.67 2.09	1.65 2.06
29	4.18 7.60	3.33 5.52	2.93 4.54	2.70 4.04	2.54 3.73	2.43 3.50	2.35 3.32	2.28 3.20	2.22 3.08	2.18 3.00	2.14 2.92	2.10 2.87	2.05 2.77	2.00 2.68	1.94 2.57	1.90 2.49	1.85 2.41	1.80 2.32	1.77 2.27	1.73 2.19	1.71 2.15	1.68 2.10	1.65 2.06	1.64 2.03
30	4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.34 3.30	2.27 3.17	2.21 3.06	2.16 2.98	2.12 2.90	2.09 2.84	2.04 2.74	1.99 2.66	1.93 2.55	1.89 2.47	1.84 2.38	1.79 2.29	1.76 2.24	1.72 2.16	1.69 2.13	1.66 2.07	1.64 2.03	1.62 2.01

(continued)

Table D. *(continued)*

0.05 (light row) and 0.01 (dark row) points for the distribution of F

Degrees of freedom for lesser mean square	Degrees of freedom for greater mean square: 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
	7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80	2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
	7.44	5.29	4.42	3.93	3.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76	2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.89	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
	7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72	2.62	2.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69	2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66	2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.90	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.06	2.09	1.92	1.88	1.82	1.78	1.75
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56	2.46	2.39	2.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.40	2.32	2.20	2.12	2.03	1.93	1.87	1.79	1.74	1.68	1.63	1.60
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47	2.37	2.30	2.18	2.09	2.00	1.90	1.84	1.76	1.71	1.64	1.60	1.56
70	3.98	3.13	2.74	2.50	2.35	2.32	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35
	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41	2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49

Degrees of freedom for lesser mean square																								
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28
	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33	2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
	6.81	4.75	3.91	3.44	3.13	2.92	2.76	2.62	2.53	2.44	2.37	2.30	2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
	6.76	4.71	3.38	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28	1.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
	6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.26	2.20	2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.38	1.28	1.19	1.11
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00
	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.00

Table D_1

Values of F exceeded by 0.025 of the values in the sampling distribution.

If, in testing the homogeneity of two sample variances, the larger variance is placed over the smaller, the number of ratios greater than any given value, equal to or greater than unity, is doubled. Therefore use of the values tabulated below, in comparing two sample variances, will provide a 0.05 level of significance.

		df for larger variance (numerator)									
		4	5	6	7	8	9	10	12	15	20
df for smaller variance (denominator)	4	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56
	5	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33
	6	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17
	7	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47
	8	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00
	9	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67
	10	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42
	12	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07
	15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76
	20	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46

Interpolation may be performed using reciprocals of the degrees of freedom.

Table E

Critical Values of *A*

For any given value of n − 1, the table shows the values of A corresponding to various levels of probability. A is significant at a given level if it is equal to or <u>less than</u> the value shown in the table.

	Level of significance for one-tailed test					
	.05	.025	.01	.005	.0005	
	Level of significance for two-tailed test					
n − 1*	.10	.05	.02	.01	.001	n − 1*
1	0.5125	0.5031	0.50049	0.50012	0.5000012	1
2	0.412	0.369	0.347	0.340	0.334	2
3	0.385	0.324	0.286	0.272	0.254	3
4	0.376	0.304	0.257	0.238	0.211	4
5	0.372	0.293	0.240	0.218	0.184	5
6	0.370	0.286	0.230	0.205	0.167	6
7	0.369	0.281	0.222	0.196	0.155	7
8	0.368	0.278	0.217	0.190	0.146	8
9	0.368	0.276	0.213	0.185	0.139	9
10	0.368	0.274	0.210	0.181	0.134	10
11	0.368	0.273	0.207	0.178	0.130	11
12	0.368	0.271	0.205	0.176	0.126	12
13	0.368	0.270	0.204	0.174	0.124	13
14	0.368	0.270	0.202	0.172	0.121	14
15	0.368	0.269	0.201	0.170	0.119	15
16	0.368	0.268	0.200	0.169	0.117	16
17	0.368	0.268	0.199	0.168	0.116	17
18	0.368	0.267	0.198	0.167	0.114	18
19	0.368	0.267	0.197	0.166	0.113	19
20	0.368	0.266	0.197	0.165	0.112	20
21	0.368	0.266	0.196	0.165	0.111	21
22	0.368	0.266	0.196	0.164	0.110	22
23	0.368	0.266	0.195	0.163	0.109	23
24	0.368	0.265	0.195	0.163	0.108	24
25	0.368	0.265	0.194	0.162	0.108	25
26	0.368	0.265	0.194	0.162	0.107	26
27	0.368	0.265	0.193	0.161	0.107	27
28	0.368	0.265	0.193	0.161	0.106	28
29	0.368	0.264	0.193	0.161	0.106	29
30	0.368	0.264	0.193	0.160	0.105	30
40	0.368	0.263	0.191	0.158	0.102	40
60	0.369	0.262	0.189	0.155	0.099	60
120	0.369	0.261	0.187	0.153	0.095	120
∞	0.370	0.260	0.185	0.151	0.092	∞

*n = number of pairs

Table F. Transformation of r to z_r.

r	z_r	r	z_r	r	z_r
.01	.010	.34	.354	.67	.811
.02	.020	.35	.366	.68	.829
.03	.030	.36	.377	.69	.848
.04	.040	.37	.389	.70	.867
.05	.050	.38	.400	.71	.887
.06	.060	.39	.412	.72	.908
.07	.070	.40	.424	.73	.929
.08	.080	.41	.436	.74	.950
.09	.090	.42	.448	.75	.973
.10	.100	.43	.460	.76	.996
.11	.110	.44	.472	.77	1.020
.12	.121	.45	.485	.78	1.045
.13	.131	.46	.497	.79	1.071
.14	.141	.47	.510	.80	1.099
.15	.151	.48	.523	.81	1.127
.16	.161	.49	.536	.82	1.157
.17	.172	.50	.549	.83	1.188
.18	.181	.51	.563	.84	1.221
.19	.192	.52	.577	.85	1.256
.20	.203	.53	.590	.86	1.293
.21	.214	.54	.604	.87	1.333
.22	.224	.55	.618	.88	1.376
.23	.234	.56	.633	.89	1.422
.24	.245	.57	.648	.90	1.472
.25	.256	.58	.663	.91	1.528
.26	.266	.59	.678	.92	1.589
.27	.277	.60	.693	.93	1.658
.28	.288	.61	.709	.94	1.738
.29	.299	.62	.725	.95	1.832
.30	.309	.63	.741	.96	1.946
.31	.321	.64	.758	.97	2.092
.32	.332	.65	.775	.98	2.298
.33	.343	.66	.793	.99	2.647

Table G. Critical Values of r_{rho} (Rank-Order Correlation Coefficient)

	Level of significance for one-tailed test			
	.05	.025	.01	.005
	Level of significance for two-tailed test			
n*	.10	.05	.02	.01
5	.900	1.000	1.000	--
6	.829	.886	.943	1.000
7	.714	.786	.893	.929
8	.643	.738	.833	.881
9	.600	.683	.783	.833
10	.564	.648	.746	.794
12	.506	.591	.712	.777
14	.456	.544	.645	.715
16	.425	.506	.601	.665
18	.399	.475	.564	.625
20	.377	.450	.534	.591
22	.359	.428	.508	.562
24	.343	.409	.485	.537
26	.329	.392	.465	.515
28	.317	.377	.448	.496
30	.306	.364	.432	.478

*n = number of pairs

Table H

Table of Functions of r

r	$\sqrt{r}$	r^2	$\sqrt{r - r^2}$	$\sqrt{1 - r}$	$1 - r^2$	$\sqrt{1 - r^2}$	$100(1 - k)$	r
						k	% Eff.	
1.00	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	100.00	1.00
.99	.9950	.9801	.0995	.1000	.0199	.1411	85.89	.99
.98	.9899	.9604	.1400	.1414	.0396	.1990	80.10	.98
.97	.9849	.9409	.1706	.1732	.0591	.2431	75.69	.97
.96	.9798	.9216	.1960	.2000	.0784	.2800	72.00	.96
.95	.9747	.9025	.2179	.2236	.0975	.3122	68.78	.95
.94	.9695	.8836	.2375	.2449	.1164	.3412	65.88	.94
.93	.9644	.8649	.2551	.2646	.1351	.3676	63.24	.93
.92	.9592	.8464	.2713	.2828	.1536	.3919	60.81	.92
.91	.9539	.8281	.2862	.3000	.1719	.4146	58.54	.91
.90	.9487	.8100	.3000	.3162	.1900	.4359	56.41	.90
.89	.9434	.7921	.3129	.3317	.2079	.4560	54.40	.89
.88	.9381	.7744	.3250	.3464	.2256	.4750	52.50	.88
.87	.9327	.7569	.3363	.3606	.2431	.4931	50.69	.87
.86	.9274	.7396	.3470	.3742	.2604	.5103	48.97	.86
.85	.9220	.7225	.3571	.3873	.2775	.5268	47.32	.85
.84	.9165	.7056	.3666	.4000	.2944	.5426	45.74	.84
.83	.9110	.6889	.3756	.4123	.3111	.5578	44.22	.83
.82	.9055	.6724	.3842	.4243	.3276	.5724	42.76	.82
.81	.9000	.6561	.3923	.4359	.3439	.5864	41.36	.81
.80	.8944	.6400	.4000	.4472	.3600	.6000	40.00	.80
.79	.8888	.6241	.4073	.4583	.3759	.6131	38.69	.79
.78	.8832	.6084	.4142	.4690	.3916	.6258	37.42	.78
.77	.8775	.5929	.4208	.4796	.4071	.6380	36.20	.77
.76	.8718	.5776	.4271	.4899	.4224	.6499	35.01	.76
.75	.8660	.5625	.4330	.5000	.4375	.6614	33.86	.75
.74	.8602	.5476	.4386	.5099	.4524	.6726	32.74	.74
.73	.8544	.5329	.4440	.5196	.4671	.6834	31.66	.73
.72	.8485	.5184	.4490	.5292	.4816	.6940	30.60	.72
.71	.8426	.5041	.4538	.5385	.4959	.7042	29.58	.71
.70	.8367	.4900	.4583	.5477	.5100	.7141	28.59	.70
.69	.8307	.4761	.4625	.5568	.5239	.7238	27.62	.69
.68	.8246	.4624	.4665	.5657	.5376	.7332	26.68	.68
.67	.8185	.4489	.4702	.5745	.5511	.7424	25.76	.67
.66	.8124	.4356	.4737	.5831	.5644	.7513	24.87	.66
.65	.8062	.4225	.4770	.5916	.5775	.7599	24.01	.65
.64	.8000	.4096	.4800	.6000	.5904	.7684	23.16	.64
.63	.7937	.3969	.4828	.6083	.6031	.7766	22.34	.63
.62	.7874	.3844	.4854	.6164	.6156	.7846	21.54	.62
.61	.7810	.3721	.4877	.6245	.6279	.7924	20.76	.61
.60	.7746	.3600	.4899	.6325	.6400	.8000	20.00	.60
.59	.7681	.3481	.4918	.6403	.6519	.8074	19.26	.59
.58	.7616	.3364	.4936	.6481	.6636	.8146	18.54	.58
.57	.7550	.3249	.4951	.6557	.6751	.8216	17.84	.57
.56	.7483	.3136	.4964	.6633	.6864	.8285	17.15	.56
.55	.7416	.3025	.4975	.6708	.6975	.8352	16.48	.55
.54	.7348	.2916	.4984	.6782	.7084	.8417	15.83	.54
.53	.7280	.2809	.4991	.6856	.7191	.8480	15.20	.53
.52	.7211	.2704	.4996	.6928	.7296	.8542	14.58	.52
.51	.7141	.2601	.4999	.7000	.7399	.8602	13.98	.51
.50	.7071	.2500	.5000	.7071	.7500	.8660	13.40	.50

Table H. *(continued)*

r	$\sqrt{r}$	r^2	$\sqrt{r - r^2}$	$\sqrt{1 - r}$	$1 - r^2$	$\sqrt{1 - r^2}$	100(1 − k)	r
						k	% Eff.	
.50	.7071	.2500	.5000	.7071	.7500	.8660	13.40	.50
.49	.7000	.2401	.4999	.7141	.7599	.8717	12.83	.49
.48	.6928	.2304	.4996	.7211	.7696	.8773	12.27	.48
.47	.6856	.2209	.4991	.7280	.7791	.8827	11.73	.47
.46	.6782	.2116	.4984	.7348	.7884	.8879	11.21	.46
.45	.6708	.2025	.4975	.7416	.7975	.8930	10.70	.45
.44	.6633	.1936	.4964	.7483	.8064	.8980	10.20	.44
.43	.6557	.1849	.4951	.7550	.8151	.9028	9.72	.43
.42	.6481	.1764	.4936	.7616	.8236	.9075	9.25	.42
.41	.6403	.1681	.4918	.7681	.8319	.9121	8.79	.41
.40	.6325	.1600	.4899	.7746	.8400	.9165	8.35	.40
.39	.6245	.1521	.4877	.7810	.8479	.9208	7.92	.39
.38	.6164	.1444	.4854	.7874	.8556	.9250	7.50	.38
.37	.6083	.1369	.4828	.7937	.8631	.9290	7.10	.37
.36	.6000	.1296	.4800	.8000	.8704	.9330	6.70	.36
.35	.5916	.1225	.4770	.8062	.8775	.9367	6.33	.35
.34	.5831	.1156	.4737	.8124	.8844	.9404	5.96	.34
.33	.5745	.1089	.4702	.8185	.8911	.9440	5.60	.33
.32	.5657	.1024	.4665	.8246	.8976	.9474	5.25	.32
.31	.5568	.0961	.4625	.8307	.9039	.9507	4.93	.31
.30	.5477	.0900	.4583	.8367	.9100	.9539	4.61	.30
.29	.5385	.0841	.4538	.8426	.9159	.9570	4.30	.29
.28	.5292	.0784	.4490	.8485	.9216	.9600	4.00	.28
.27	.5196	.0729	.4440	.8544	.9271	.9629	3.71	.27
.26	.5099	.0676	.4386	.8602	.9324	.9656	3.44	.26
.25	.5000	.0625	.4330	.8660	.9375	.9682	3.18	.25
.24	.4899	.0576	.4271	.8718	.9424	.9708	2.92	.24
.23	.4796	.0529	.4208	.8775	.9471	.9732	2.68	.23
.22	.4690	.0484	.4142	.8832	.9516	.9755	2.45	.22
.21	.4583	.0441	.4073	.8888	.9559	.9777	2.23	.21
.20	.4472	.0400	.4000	.8944	.9600	.9798	2.02	.20
.19	.4359	.0361	.3923	.9000	.9639	.9818	1.82	.19
.18	.4243	.0324	.3842	.9055	.9676	.9837	1.63	.18
.17	.4123	.0289	.3756	.9110	.9711	.9854	1.46	.17
.16	.4000	.0256	.3666	.9165	.9744	.9871	1.29	.16
.15	.3873	.0225	.3571	.9220	.9775	.9887	1.13	.15
.14	.3742	.0196	.3470	.9274	.9804	.9902	.98	.14
.13	.3606	.0169	.3363	.9327	.9831	.9915	.85	.13
.12	.3464	.0144	.3250	.9381	.9856	.9928	.72	.12
.11	.3317	.0121	.3129	.9434	.9879	.9939	.61	.11
.10	.3162	.0100	.3000	.9487	.9900	.9950	.50	.10
.09	.3000	.0081	.2862	.9539	.9919	.9959	.41	.09
.08	.2828	.0064	.2713	.9592	.9936	.9968	.32	.08
.07	.2646	.0049	.2551	.9644	.9951	.9975	.25	.07
.06	.2449	.0036	.2375	.9695	.9964	.9982	.18	.06
.05	.2236	.0025	.2179	.9747	.9975	.9987	.13	.05
.04	.2000	.0016	.1960	.9798	.9984	.9992	.08	.04
.03	.1732	.0009	.1706	.9849	.9991	.9995	.05	.03
.02	.1414	.0004	.1400	.9899	.9996	.9998	.02	.02
.01	.1000	.0001	.0995	.9950	.9999	.9999	.01	.01
.00	.0000	.0000	.0000	1.0000	1.0000	1.0000	.00	.00

Table I_1

Critical Values of U and U' for a One-tailed Test at $\alpha = 0.005$ or a Two-tailed Test at $\alpha = 0.01$

To be significant for any given n_1 and n_2: Obtained U must be equal to or less than the value shown in the table. Obtained U′ must be equal to or greater than the value shown in the table.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0 38	0 40
3	--	--	--	--	--	--	--	--	0 27	0 30	0 33	1 35	1 38	1 41	2 43	2 46	2 49	2 52	3 54	3 57
4	--	--	--	--	--	0 24	0 28	1 31	1 35	2 38	2 42	3 45	3 49	4 52	5 55	5 59	6 62	6 66	7 69	8 72
5	--	--	--	--	0 25	1 29	1 34	2 38	3 42	4 46	5 50	6 54	7 58	7 63	8 67	9 71	10 75	11 79	12 83	13 87
6	--	--	--	0 24	1 29	2 34	3 39	4 44	5 49	6 54	7 59	9 63	10 68	11 73	12 78	13 83	15 87	16 92	17 97	18 102
7	--	--	--	0 28	1 34	3 39	4 45	6 50	7 56	9 61	10 67	12 72	13 78	15 83	16 89	18 94	19 100	21 105	22 111	24 116
8	--	--	--	1 31	2 38	4 44	6 50	7 57	9 63	11 69	13 75	15 81	17 87	18 94	20 100	22 106	24 112	26 118	28 124	30 130
9	--	--	0 27	1 35	3 42	5 49	7 56	9 63	11 70	13 77	16 83	18 90	20 97	22 104	24 111	27 117	29 124	31 131	33 138	36 144
10	--	--	0 30	2 38	4 46	6 54	9 61	11 69	13 77	16 84	18 92	21 99	24 106	26 114	29 121	31 129	34 136	37 143	39 151	42 158
11	--	--	0 33	2 42	5 50	7 59	10 67	13 75	16 83	18 92	21 100	24 108	27 116	30 124	33 132	36 140	39 148	42 156	45 164	48 172
12	--	--	1 35	3 45	6 54	9 63	12 72	15 81	18 90	21 99	24 108	27 117	31 125	34 134	37 143	41 151	44 160	47 169	51 177	54 186
13	--	--	1 38	3 49	7 58	10 68	13 78	17 87	20 97	24 106	27 116	31 125	34 125	38 144	42 153	45 163	49 172	53 181	56 191	60 200
14	--	--	1 41	4 52	7 63	11 73	15 83	18 94	22 104	26 114	30 124	34 134	38 144	42 154	46 164	50 174	54 184	58 194	63 203	67 213
15	--	--	2 43	5 55	8 67	12 78	16 89	20 100	24 111	29 121	33 132	37 143	42 153	46 164	51 174	55 185	60 195	64 206	69 216	73 227
16	--	--	2 46	5 59	9 71	13 83	18 94	22 106	27 117	31 129	36 140	41 151	45 163	50 174	55 185	60 196	65 207	70 218	74 230	79 241
17	--	--	2 49	6 62	10 75	15 87	19 100	24 112	29 124	34 148	39 148	44 160	49 172	54 184	60 195	65 207	70 219	75 231	81 242	86 254
18	--	--	2 52	6 66	11 79	16 92	21 105	26 118	31 131	37 143	42 156	47 169	53 181	58 194	64 206	70 218	75 231	81 243	87 255	92 268
19	--	0 38	3 54	7 69	12 83	17 97	22 111	28 124	33 138	39 151	45 164	51 177	56 191	63 203	69 216	74 230	81 242	87 255	93 268	99 281
20	--	0 40	3 57	8 72	13 87	18 102	24 116	30 130	36 144	42 158	48 172	54 186	60 200	67 213	73 227	79 241	86 254	92 268	99 281	105 295

(Dashes in the body of the table indicate that no decision is possible at the stated level of significance.)

Table I_2

Critical Values of U and U' for a One-tailed Test at $\alpha = 0.01$ or a Two-tailed Test at $\alpha = 0.02$

To be significant for any given n_1 and n_2: Obtained U must be equal to or less than the value shown in the table. Obtained U′ must be equal to or greater than the value shown in the table.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	--	--	--	--	--	0 26	0 28	0 30	0 32	0 34	0 36	1 37	1 39
3	--	--	--	--	--	--	0 21	0 24	1 26	1 29	1 32	2 34	2 37	2 40	3 42	3 45	4 47	4 50	4 52	5 55
4	--	--	--	--	0 20	1 23	1 27	2 30	3 33	3 37	4 40	5 43	5 47	6 50	7 53	7 57	8 60	9 63	9 67	10 70
5	--	--	--	0 20	1 24	2 28	3 32	4 36	5 40	6 44	7 48	8 52	9 56	10 60	11 64	12 68	13 72	14 76	15 80	16 84
6	--	--	--	1 23	2 28	3 33	4 38	6 42	7 47	8 52	9 57	11 61	12 66	13 71	15 75	16 80	18 84	19 89	20 94	22 98
7	--	--	0 21	1 27	3 32	4 38	6 43	7 49	9 54	11 59	12 65	14 70	16 75	17 81	19 86	21 91	23 96	24 102	26 107	28 112
8	--	--	0 24	2 30	4 36	6 42	7 49	9 55	11 61	13 67	15 73	17 79	20 84	22 90	24 96	26 102	28 108	30 114	32 120	34 126
9	--	--	1 26	3 33	5 40	7 47	9 54	11 61	14 67	16 74	18 81	21 87	23 94	26 100	28 107	31 113	33 120	36 126	38 133	40 140
10	--	--	1 29	3 37	6 44	8 52	11 59	13 67	16 74	19 81	22 88	24 96	27 103	30 110	33 117	36 124	38 132	41 139	44 146	47 153
11	--	--	1 32	4 40	7 48	9 57	12 65	15 73	18 81	22 88	25 96	28 104	31 112	34 120	37 128	41 135	44 143	47 151	50 159	53 167
12	--	--	2 34	5 43	8 52	11 61	14 70	17 79	21 87	24 96	28 104	31 113	35 121	38 130	42 138	46 146	49 155	53 163	56 172	60 180
13	--	0 26	2 37	5 47	9 56	12 66	16 75	20 84	23 94	27 103	31 112	35 121	39 130	43 139	47 148	51 157	55 166	59 175	63 184	67 193
14	--	0 28	2 40	6 50	10 60	13 71	17 81	22 90	26 100	30 110	34 120	38 130	43 139	47 149	51 159	56 168	60 178	65 187	69 197	73 207
15	--	0 30	3 42	7 53	11 64	15 75	19 86	24 96	28 107	33 117	37 128	42 138	47 148	51 159	56 169	61 179	66 189	70 200	75 210	80 220
16	--	0 32	3 45	7 57	12 68	16 80	21 91	26 102	31 113	36 124	41 135	46 146	51 157	56 168	61 179	66 190	71 201	76 212	82 222	87 233
17	--	0 34	4 47	8 60	13 72	18 84	23 96	28 108	33 120	38 132	44 143	49 155	55 166	60 178	66 189	71 201	77 212	82 224	88 234	93 247
18	--	0 36	4 50	9 63	14 76	19 89	24 102	30 114	36 126	41 139	47 151	53 163	59 175	65 187	70 200	76 212	82 224	88 236	94 248	100 260
19	--	1 37	4 53	9 67	15 80	20 94	26 107	32 120	38 133	44 146	50 159	56 172	63 184	69 197	75 210	82 222	88 235	94 248	101 260	107 273
20	--	1 39	5 55	10 70	16 84	22 98	28 112	34 126	40 140	47 153	53 167	60 180	67 193	73 207	80 220	87 233	93 247	100 260	107 273	114 286

(Dashes in the body of the table indicate that no decision is possible at the stated level of significance.)

Table I$_3$

Critical Values of U and U' for a One-tailed Test at $\alpha = 0.025$ or a Two-tailed Test at $\alpha = 0.05$

To be significant for any given n_1 and n_2: Obtained U must be equal to or less than the value shown in the table. Obtained U′ must be equal to or greater than the value shown in the table.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	0 / 16	0 / 18	0 / 20	0 / 22	1 / 23	1 / 25	1 / 27	1 / 29	1 / 31	2 / 32	2 / 34	2 / 36	2 / 38
3	--	--	--	--	0 / 15	1 / 17	1 / 20	2 / 22	2 / 25	3 / 27	3 / 30	4 / 32	4 / 35	5 / 37	5 / 40	6 / 42	6 / 45	7 / 47	7 / 50	8 / 52
4	--	--	--	0 / 16	1 / 19	2 / 22	3 / 25	4 / 28	4 / 32	5 / 35	6 / 38	7 / 41	8 / 44	9 / 47	10 / 50	11 / 53	11 / 57	12 / 60	13 / 63	13 / 67
5	--	--	0 / 15	1 / 19	2 / 23	3 / 27	5 / 30	6 / 34	7 / 38	8 / 42	9 / 46	11 / 49	12 / 53	13 / 57	14 / 61	15 / 65	17 / 68	18 / 72	19 / 76	20 / 80
6	--	--	1 / 17	2 / 22	3 / 27	5 / 31	6 / 36	8 / 40	10 / 44	11 / 49	13 / 53	14 / 58	16 / 62	17 / 67	19 / 71	21 / 75	22 / 80	24 / 84	25 / 89	27 / 93
7	--	--	1 / 20	3 / 25	5 / 30	6 / 36	8 / 41	10 / 46	12 / 51	14 / 56	16 / 61	18 / 66	20 / 71	22 / 76	24 / 81	26 / 86	28 / 91	30 / 96	32 / 101	34 / 106
8	--	0 / 16	2 / 22	4 / 28	6 / 34	8 / 40	10 / 46	13 / 51	15 / 57	17 / 63	19 / 69	22 / 74	24 / 80	26 / 86	29 / 91	31 / 97	34 / 102	36 / 108	38 / 111	41 / 119
9	--	0 / 18	2 / 25	4 / 32	7 / 38	10 / 44	12 / 51	15 / 57	17 / 64	20 / 70	23 / 76	26 / 82	28 / 89	31 / 95	34 / 101	37 / 107	39 / 114	42 / 120	45 / 126	48 / 132
10	--	0 / 20	3 / 27	5 / 35	8 / 42	11 / 49	14 / 56	17 / 63	20 / 70	23 / 77	26 / 84	29 / 91	33 / 97	36 / 104	39 / 111	42 / 118	45 / 125	48 / 132	52 / 138	55 / 145
11	--	0 / 22	3 / 30	6 / 38	9 / 46	13 / 53	16 / 61	19 / 69	23 / 76	26 / 84	30 / 91	33 / 99	37 / 106	40 / 114	44 / 121	47 / 129	51 / 136	55 / 143	58 / 151	62 / 158
12	--	1 / 23	4 / 32	7 / 41	11 / 49	14 / 58	18 / 66	22 / 74	26 / 82	29 / 91	33 / 99	37 / 107	41 / 115	45 / 123	49 / 131	53 / 139	57 / 147	61 / 155	65 / 163	69 / 171
13	--	1 / 25	4 / 35	8 / 44	12 / 53	16 / 62	20 / 71	24 / 80	28 / 89	33 / 97	37 / 106	41 / 115	45 / 124	50 / 132	54 / 141	59 / 149	63 / 158	67 / 167	72 / 175	76 / 184
14	--	1 / 27	5 / 37	9 / 47	13 / 51	17 / 67	22 / 76	26 / 86	31 / 95	36 / 104	40 / 114	45 / 123	50 / 132	55 / 141	59 / 151	64 / 160	67 / 171	74 / 178	78 / 188	83 / 197
15	--	1 / 29	5 / 40	10 / 50	14 / 61	19 / 71	24 / 81	29 / 91	34 / 101	39 / 111	44 / 121	49 / 131	54 / 141	59 / 151	64 / 161	70 / 170	75 / 180	80 / 190	85 / 200	90 / 210
16	--	1 / 31	6 / 42	11 / 53	15 / 65	21 / 75	26 / 86	31 / 97	37 / 107	42 / 118	47 / 129	53 / 139	59 / 149	64 / 160	70 / 170	75 / 181	81 / 191	86 / 202	92 / 212	98 / 222
17	--	2 / 32	6 / 45	11 / 57	17 / 68	22 / 80	28 / 91	34 / 102	39 / 114	45 / 125	51 / 136	57 / 147	63 / 158	67 / 171	75 / 180	81 / 191	87 / 202	93 / 213	99 / 224	105 / 235
18	--	2 / 34	7 / 47	12 / 60	18 / 72	24 / 84	30 / 96	36 / 108	42 / 120	48 / 132	55 / 143	61 / 155	67 / 167	74 / 178	80 / 190	86 / 202	93 / 213	99 / 225	106 / 236	112 / 248
19	--	2 / 36	7 / 50	13 / 63	19 / 76	25 / 89	32 / 101	38 / 114	45 / 126	52 / 138	58 / 151	65 / 163	72 / 175	78 / 188	85 / 200	92 / 212	99 / 224	106 / 236	113 / 248	119 / 261
20	--	2 / 38	8 / 52	13 / 67	20 / 80	27 / 93	34 / 106	41 / 119	48 / 132	55 / 145	62 / 158	69 / 171	76 / 184	83 / 197	90 / 210	98 / 222	105 / 235	112 / 248	119 / 261	127 / 273

(Dashes in the body of the table indicate that no decision is possible at the stated level of significance.)

Table I₄

Critical Values of U and U' for a One-tailed Test at $\alpha = 0.05$ or a Two-tailed Test at $\alpha = 0.10$

To be significant for any given n_1 and n_2: Obtained U must be equal to or less than the value shown in the table. Obtained U′ must be equal to or greater than the value shown in the table.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0 19	0 20
2	--	--	--	--	0 10	0 12	0 14	1 15	1 17	1 19	1 21	2 22	2 24	2 26	3 27	3 29	3 31	4 32	4 34	4 36
3	--	--	0 9	0 12	1 14	2 16	2 19	3 21	3 24	4 26	5 28	5 31	6 33	7 35	7 38	8 40	9 42	9 45	10 47	11 49
4	--	--	0 12	1 15	2 18	3 21	4 24	5 27	6 30	7 33	8 36	9 39	10 42	11 45	12 48	14 50	15 53	16 56	17 59	18 62
5	--	0 10	1 14	2 18	4 21	5 25	6 29	8 32	9 36	11 39	12 43	13 47	15 50	16 54	18 57	19 61	20 65	22 68	23 72	25 75
6	--	0 12	2 16	3 21	5 25	7 29	8 34	10 38	12 42	14 46	16 50	17 55	19 59	21 63	23 67	25 71	26 76	28 80	30 84	32 88
7	--	0 14	2 19	4 24	6 29	8 34	11 38	13 43	15 48	17 53	19 58	21 63	24 67	26 72	28 77	30 82	33 86	35 91	37 96	39 101
8	--	1 15	3 21	5 27	8 32	10 38	13 43	15 49	18 54	20 60	23 65	26 70	28 76	31 81	33 87	36 92	39 97	41 103	44 108	47 113
9	--	1 17	3 24	6 30	9 36	12 42	15 48	18 54	21 60	24 66	27 72	30 78	33 84	36 90	39 96	42 102	45 108	48 114	51 120	54 126
10	--	1 19	4 26	7 33	11 39	14 46	17 53	20 60	24 66	27 73	31 79	34 86	37 93	41 99	44 106	48 112	51 119	55 125	58 132	62 138
11	--	1 21	5 28	8 36	12 43	16 50	19 58	23 65	27 72	31 79	34 87	38 94	42 101	46 108	50 115	54 122	57 130	61 137	65 144	69 151
12	--	2 22	5 31	9 39	13 47	17 55	21 63	26 70	30 78	34 86	38 94	42 102	47 109	51 117	55 125	60 132	64 140	68 148	72 156	77 163
13	--	2 24	6 33	10 42	15 50	19 59	24 67	28 76	33 84	37 93	42 101	47 109	51 118	56 126	61 134	65 143	70 151	75 159	80 167	84 176
14	--	2 26	7 35	11 45	16 54	21 63	26 72	31 81	36 90	41 99	46 108	51 117	56 126	61 135	66 144	71 153	77 161	82 170	87 179	92 188
15	--	3 27	7 38	12 48	18 57	23 67	28 77	33 87	39 96	44 106	50 115	55 125	61 134	66 144	72 153	77 163	83 172	88 182	94 191	100 200
16	--	3 29	8 40	14 50	19 61	25 71	30 82	36 92	42 102	48 112	54 122	60 132	65 143	71 153	77 163	83 173	89 183	95 193	101 203	107 213
17	--	3 31	9 42	15 53	20 65	26 76	33 86	39 97	45 108	51 119	57 130	64 140	70 151	77 161	83 172	89 183	96 193	102 204	109 214	115 225
18	--	4 32	9 45	16 56	22 68	28 80	35 91	41 103	48 114	55 123	61 137	68 148	75 159	82 170	88 182	95 193	102 204	109 215	116 226	123 237
19	0 19	4 34	10 47	17 59	23 72	30 84	37 96	44 108	51 120	58 132	65 144	72 156	80 167	87 179	94 191	101 203	109 214	116 226	123 238	130 250
20	0 20	4 36	11 49	18 62	25 75	32 88	39 101	47 113	54 126	62 138	69 151	77 163	84 176	92 188	100 200	107 213	115 225	123 237	130 250	138 262

(Dashes in the body of the table indicate that no decision is possible at the stated level of significance.)

Table J

Critical Values of *T* at Various Levels of Probability

The symbol T denotes the smaller sum of ranks associated with differences that are all of the same sign. For any given N (number of ranked differences), the obtained T is significant at a given level if it is equal to or less than the value shown in the table.

	Level of significance for one-tailed test			
	.05	.025	.01	.005
	Level of significance for two-tailed test			
N	.10	.05	.02	.01
5	0	--	--	--
6	2	0	--	--
7	3	2	0	--
8	5	3	1	0
9	8	5	3	1
10	10	8	5	3
11	13	10	7	5
12	17	13	9	7
13	21	17	12	9
14	25	21	15	12
15	30	25	19	15
16	35	29	23	19
17	41	34	27	23
18	47	40	32	27
19	53	46	37	32
20	60	52	43	37
21	67	58	49	42
22	75	65	55	48
23	83	73	62	54
24	91	81	69	61
25	100	89	76	68
26	110	98	84	75
27	119	107	92	83

	Level of significance for one-tailed test			
	.05	.025	.01	.005
	Level of significance for two-tailed test			
N	.10	.05	.02	.01
28	130	116	101	91
29	140	126	110	100
30	151	137	120	109
31	163	147	130	118
32	175	159	140	128
33	187	170	151	138
34	200	182	162	148
35	213	195	173	159
36	227	208	185	171
37	241	221	198	182
38	256	235	211	194
39	271	249	224	207
40	286	264	238	220
41	302	279	252	233
42	319	294	266	247
43	336	310	281	261
44	353	327	296	276
45	371	343	312	291
46	389	361	328	307
47	407	378	345	322
48	426	396	362	339
49	446	415	379	355
50	466	434	397	373

(Slight discrepancies will be found between the critical values appearing in the table above and in Table 2 of the 1964 revision of F. Wilcoxon, and R.A. Wilcox, Some Rapid Approximate Statistical Procedures, New York, Lederle Laboratories, 1964. The disparity reflects the latter's policy of selecting the critical value nearest a given significance level, occasionally overstepping that level. For example, for N = 8,

the probability of a T of 3 = 0.0390 (two-tail)

and

the probability of a T of 4 = 0.0546 (two-tail).

Wilcoxon and Wilcox select a T of 4 as the critical value at the 0.05 level of significance (two-tail), whereas Table J reflects a more conservative policy by setting a T of 3 as the critical value at this level.)

Table K

Factorials of Numbers 1 to 20

N	N!
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5040
8	40320
9	362880
10	3628800
11	39916800
12	479001600
13	6227020800
14	87178291200
15	1307674368000
16	20922789888000
17	355687428096000
18	6402373705728000
19	121645100408832000
20	2432902008176640000

Table L

Binomial Coefficients

N	$\binom{N}{0}$	$\binom{N}{1}$	$\binom{N}{2}$	$\binom{N}{3}$	$\binom{N}{4}$	$\binom{N}{5}$	$\binom{N}{6}$	$\binom{N}{7}$	$\binom{N}{8}$	$\binom{N}{9}$	$\binom{N}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

Table M

Cumulative Probabilities Associated with Values As Small As Observed Values of x in the Binomial Test (x is the smaller of the observed frequencies)

Given in the body of this table are one-tailed probabilities under H_0 for the binomial test when $P = Q = 1/2$. To save space, decimal points are omitted in the p's. (x = smaller of the observed frequencies)

N \ x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	031	188	500	812	969	†										
6	016	109	344	656	891	984	†									
7	008	062	227	500	773	938	992	†								
8	004	035	145	363	637	855	965	996	†							
9	002	020	090	254	500	746	910	980	998	†						
10	001	011	055	172	377	623	828	945	989	999	†					
11		006	033	113	274	500	726	887	967	994	†	†				
12		003	019	073	194	387	613	806	927	981	997	†	†			
13		002	011	046	133	291	500	709	867	954	989	998	†	†		
14		001	006	029	090	212	395	605	788	910	971	994	999	†	†	
15			004	018	059	151	304	500	696	849	941	982	996	†	†	†
16			002	011	038	105	227	402	598	773	895	962	989	998	†	†
17			001	006	025	072	166	315	500	685	834	928	975	994	999	†
18			001	004	015	048	119	240	407	593	760	881	952	985	996	999
19				002	010	032	084	180	324	500	676	820	916	968	990	998
20				001	006	021	058	132	252	412	588	748	868	942	979	994
21				001	004	013	039	095	192	332	500	668	808	905	961	987
22					002	008	026	067	143	262	416	584	738	857	933	974
23					001	005	017	047	105	202	339	500	661	798	895	953
24					001	003	011	032	076	154	271	419	581	729	846	924
25						002	007	022	054	115	212	345	500	655	788	885

† 1.0 or approximately 1.0.

Table N

Critical Values of x at the 0.05 (lightface type) and 0.01 (boldface type) Levels of Significance for Various Values of P and Q when $N \leq 10$

x = frequency in the category with P probability of occurrence. The obtained x must be equal to or greater than the value shown in the table for significance at the chosen level.

N	P .01 Q .99	.05 .95	.10 .90	.20 .80	.25 .75	.30 .70	1/3 2/3	.40 ..60
2	1 **2**	2 **2**	2 **2**	2 --	-- --	-- --	-- --	-- --
3	1 **2**	2 **2**	2 **3**	3 **3**	3 --	3 --	3 --	-- --
4	1 **2**	2 **3**	2 **3**	3 **4**	4 **4**	4 **4**	4 --	4 --
5	1 **2**	2 **3**	3 **3**	4 **4**	4 **5**	4 **5**	4 **5**	5 --
6	2 **2**	2 **3**	3 **4**	4 **5**	4 **5**	5 **6**	5 **6**	5 **6**
7	2 **2**	2 **3**	3 **4**	4 **5**	5 **6**	5 **6**	5 **6**	6 **7**
8	2 **2**	3 **3**	3 **4**	5 **6**	5 **6**	6 **7**	6 **7**	6 **7**
9	2 **2**	3 **3**	4 **4**	5 **6**	5 **6**	6 **7**	6 **7**	7 **8**
10	2 **2**	3 **4**	4 **5**	5 **6**	6 **7**	6 **8**	7 **8**	8 **9**

(Dashes in the body of the table indicate that no decision is possible at the stated level of significance.)

Table O

Some Approximately Normal Populations for Sampling Experiments

X	f	f	f	f
2	4	2	0	0
3	54	27	6	3
4	242	121	24	12
5	400	200	40	20
6	242	121	24	12
7	54	27	6	3
8	4	2	0	0
N	1000	500	100	50
μ	5.00	5.00	5.00	5.00
σ	.99	.99	.98	.98

Table P. Percentage Points of the Studentized Range

Error df	α	k = *number of means or number of steps between ordered means*									
		2	3	4	5	6	7	8	9	10	11
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
	.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
	.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61
	.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13
12	.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51
	.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31
	.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
	.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46
17	.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19
24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01
	.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02
30	.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92
	.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82
	.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73
	.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53
120	.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37
∞	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23

Table Q. Squares, Square Roots, and Reciprocals of Numbers from 1 to 1000

N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N
1	1	1.0000	1.000000	61	3721	7.8102	.016393	121	14641	11.0000	.00826446
2	4	1.4142	.500000	62	3844	7.8740	.016129	122	14884	11.0454	.00819672
3	9	1.7321	.333333	63	3969	7.9373	.015873	123	15129	11.0905	.00813008
4	16	2.0000	.250000	64	4096	8.0000	.015625	124	15376	11.1355	.00800452
5	25	2.2361	.200000	65	4225	8.0623	.015385	125	15625	11.1803	.00800000
6	36	2.4495	.166667	66	4356	8.1240	.015152	126	15876	11.2250	.00793651
7	49	2.6458	.142857	67	4489	8.1854	.014925	127	16129	11.2694	.00787402
8	64	2.8284	.125000	68	4624	8.2462	.014706	128	16384	11.3137	.00781250
9	81	3.0000	.111111	69	4761	8.3066	.014493	129	16641	11.3578	.00775194
10	100	3.1623	.100000	70	4900	8.3666	.014286	130	16900	11.4018	.00769231
11	121	3.3166	.090909	71	5041	8.4261	.014085	131	17161	11.4455	.00763359
12	144	3.4641	.083333	72	5184	8.4853	.013889	132	17424	11.4891	.00757576
13	169	3.6056	.076923	73	5329	8.5440	.013699	133	17689	11.5326	.00751880
14	196	3.7417	.071429	74	5476	8.6023	.013514	134	17956	11.5758	.00746269
15	225	3.8730	.066667	75	5625	8.6603	.013333	135	18225	11.6190	.00740741
16	256	4.0000	.062500	76	5776	8.7178	.013158	136	18496	11.6619	.00735294
17	289	4.1231	.058824	77	5929	8.7750	.012987	137	18769	11.7047	.00729927
18	324	4.2426	.055556	78	6084	8.8318	.012821	138	19044	11.7473	.00724638
19	361	4.3589	.052632	79	6241	8.8882	.012658	139	19321	11.7898	.00719424
20	400	4.4721	.050000	80	6400	8.9443	.012500	140	19600	11.8322	.00714286
21	441	4.5826	.047619	81	6561	9.0000	.012346	141	19881	11.8743	.00709220
22	484	4.6904	.045455	82	6724	9.0554	.012195	142	20164	11.9164	.00704225
23	529	4.7958	.043478	83	6889	9.1104	.012048	143	20449	11.9583	.00699301
24	576	4.8990	.041667	84	7056	9.1652	.011905	144	20736	12.0000	.00694444
25	625	5.0000	.040000	85	7225	9.2195	.011765	145	21025	12.0416	.00689655
26	676	5.0990	.038462	86	7396	9.2736	.011628	146	21316	12.0830	.00684932
27	729	5.1962	.037037	87	7569	9.3274	.011494	147	21609	12.1244	.00680272
28	784	5.2915	.035714	88	7744	9.3808	.011364	148	21904	12.1655	.00675676
29	841	5.3852	.034483	89	7921	9.4340	.011236	149	22201	12.2066	.00671141
30	900	5.4772	.033333	90	8100	9.4868	.011111	150	22500	12.2474	.00666667
31	961	5.5678	.032258	91	8281	9.5394	.010989	151	22801	12.2882	.00662252
32	1024	5.6569	.031250	92	8464	9.5917	.010870	152	23104	12.3288	.00657895
33	1089	5.7446	.030303	93	8649	9.6437	.010753	153	23409	12.3693	.00653595
34	1156	5.8310	.029412	94	8836	9.6954	.010638	154	23716	12.4097	.00649351
35	1225	5.9161	.028571	95	9025	9.7468	.010526	155	24025	12.4499	.00645161
36	1296	6.0000	.027778	96	9216	9.7980	.010417	156	24336	12.4900	.00641026
37	1369	6.0828	.027027	97	9409	9.8489	.010309	157	24649	12.5300	.00636943
38	1444	6.1644	.026316	98	9604	9.8995	.010204	158	24964	12.5698	.00632911
39	1521	6.2450	.025641	99	9801	9.9499	.010101	159	25281	12.6095	.00628931
40	1600	6.3246	.025000	100	10000	10.0000	.010000	160	25600	12.6491	.00625000
41	1681	6.4031	.024390	101	10201	10.0499	.00990099	161	25921	12.6886	.00621118
42	1764	6.4807	.023810	102	10404	10.0995	.00980392	162	26244	12.7279	.00617284
43	1849	6.5574	.023256	103	10609	10.1489	.00970874	163	26569	12.7671	.00613497
44	1936	6.6332	.022727	104	10816	10.1980	.00961538	164	26896	12.8062	.00609756
45	2025	6.7082	.022222	105	11025	10.2470	.00952381	165	27225	12.8452	.00606061
46	2116	6.7823	.021739	106	11236	10.2956	.00943396	166	27556	12.8841	.00602410
47	2209	6.8557	.021277	107	11449	10.3441	.00934579	167	27889	12.9228	.00598802
48	2304	6.9282	.020833	108	11664	10.3923	.00925926	168	28224	12.9615	.00595238
49	2401	7.0000	.020408	109	11881	10.4403	.00917431	169	28561	13.0000	.00591716
50	2500	7.0711	.020000	110	12100	10.4881	.00909091	170	28900	13.0384	.00588235
51	2601	7.1414	.019608	111	12321	10.5357	.00900901	171	29241	13.0767	.00584795
52	2704	7.2111	.019231	112	12544	10.5830	.00892857	172	29584	13.1149	.00581395
53	2809	7.2801	.018868	113	12769	10.6301	.00884956	173	29929	13.1529	.00578035
54	2916	7.3485	.018519	114	12996	10.6771	.00877193	174	30276	13.1909	.00574713
55	3025	7.4162	.018182	115	13225	10.7238	.00869565	175	30625	13.2288	.00571429
56	3136	7.4833	.017857	116	13456	10.7703	.00862069	176	30976	13.2665	.00568182
57	3249	7.5498	.017544	117	13689	10.8167	.00854701	177	31329	13.3041	.00564972
58	3364	7.6158	.017241	118	13924	10.8628	.00847458	178	31684	13.3417	.00561798
59	3481	7.6811	.016949	119	14161	10.9087	.00840336	179	32041	13.3791	.00558659
60	3600	7.7460	.016667	120	14400	10.9545	.00833333	180	32400	13.4164	.00555556

Table Q. *(continued)*

N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N
181	32761	13.4536	.00552486	241	58081	15.5242	.00414938	301	90601	17.3494	.00332226
182	33124	13.4907	.00549451	242	58564	15.5563	.00413223	302	91204	17.3781	.00331126
183	33489	13.5277	.00546448	243	59049	15.5885	.00411523	303	91809	17.4069	.00330033
184	33856	13.5647	.00543478	244	59536	15.6205	.00409836	304	92416	17.4356	.00328047
185	34225	13.6015	.00540541	245	60025	15.6525	.00408163	305	93025	17.4642	.00328947
186	34596	13.6382	.00537634	246	60516	15.6844	.00406504	306	93636	17.4929	.00326797
187	34969	13.6748	.00534759	247	61009	15.7162	.00404858	307	94249	17.5214	.00325733
188	35344	13.7113	.00531915	248	61504	15.7480	.00403226	308	94864	17.5499	.00321675
189	35721	13.7477	.00529101	249	62001	15.7797	.00401606	309	95481	17.5784	.00323625
190	36100	13.7840	.00526316	250	62500	15.8114	.00400000	310	96100	17.6068	.00322581
191	36481	13.8203	.00523560	251	63001	15.8430	.00398406	311	96721	17.6352	.00321543
192	36864	13.8564	.00520833	252	63504	15.8745	.00396825	312	97344	17.6635	.00320513
193	37249	13.8924	.00518135	253	64009	15.9060	.00395257	313	97969	17.6918	.00319489
194	37636	13.9284	.00515464	254	64516	15.9374	.00393701	314	98596	17.7200	.00318471
195	38025	13.9642	.00512821	255	65025	15.9687	.00392157	315	99225	17.7482	.00317460
196	38416	14.0000	.00510204	256	65536	16.0000	.00390625	316	99856	17.7764	.00316456
197	38809	14.0357	.00507614	257	66049	16.0312	.00389105	317	100489	17.8045	.00315457
198	39204	14.0712	.00505051	258	66564	16.0624	.00387597	318	101124	17.8326	.00314465
199	39601	14.1067	.00502513	259	67081	16.0935	.00386100	319	101761	17.8606	.00313480
200	40000	14.1421	.00500000	260	67600	16.1245	.00384615	320	102400	17.8885	.00312500
201	40401	14.1774	.00497512	261	68121	16.1555	.00383142	321	103041	17.9165	.00311526
202	40804	14.2127	.00495050	262	68644	16.1864	.00381679	322	103684	17.9444	.00310559
203	41209	14.2478	.00492611	263	69169	16.2173	.00380228	323	104329	17.9722	.00309598
204	41616	14.2829	.00490196	264	69696	16.2481	.00378788	324	104976	18.0000	.00308642
205	42025	14.3178	.00487805	265	70225	16.2788	.00377358	325	105625	18.0278	.00307692
206	42436	14.3527	.00485437	266	70756	16.3095	.00375940	326	106276	18.0555	.00306748
207	42849	14.3875	.00483092	267	71289	16.3401	.00374532	327	106929	18.0831	.00305810
208	43264	14.4222	.00480769	268	71824	16.3707	.00373134	328	107584	18.1108	.00304878
209	43681	14.4568	.00478469	269	72361	16.4012	.00371747	329	108241	18.1384	.00303951
210	44100	14.4914	.00476190	270	72900	16.4317	.00370370	330	108900	18.1659	.00303030
211	44521	14.5258	.00473934	271	73441	16.4621	.00369004	331	109561	18.1934	.00302115
212	44944	14.5602	.00471698	272	73984	16.4924	.00367647	332	110224	18.2209	.00301205
213	45369	14.5945	.00469484	273	74529	16.5227	.00366300	333	110889	18.2483	.00300300
214	45796	14.6287	.00467290	274	75076	16.5529	.00364964	334	111556	18.2757	.00299401
215	46225	14.6629	.00465116	275	75625	16.5831	.00363636	335	112225	18.3030	.00298507
216	46656	14.6969	.00462963	276	76176	16.6132	.00362319	336	112896	18.3303	.00297619
217	47089	14.7309	.00460829	277	76729	16.6433	.00361011	337	113569	18.3576	.00296736
218	47524	14.7648	.00458716	278	77284	16.6733	.00359712	338	114244	18.3848	.00295858
219	47961	14.7986	.00456621	279	77841	16.7033	.00358423	339	114921	18.4120	.00294985
220	48400	14.8324	.00454545	280	78400	16.7332	.00357143	340	115600	18.4391	.00294118
221	48841	14.8661	.00452489	281	78961	16.7631	.00355872	341	116281	18.4662	.00293255
222	49284	14.8997	.00450450	282	79524	16.7929	.00354610	342	116964	18.4932	.00292398
223	49729	14.9332	.00448430	283	80089	16.8226	.00353357	343	117649	18.5203	.00291545
224	50176	14.9666	.00446429	284	80656	16.8523	.00352113	344	118336	18.5472	.00290698
225	50625	15.0000	.00444444	285	81225	16.8819	.00350877	345	119025	18.5742	.00289855
226	51076	15.0333	.00442478	286	81796	16.9115	.00349650	346	119716	18.6011	.00289017
227	51529	15.0665	.00440529	287	82369	16.9411	.00348432	347	120409	18.6279	.00288184
228	51984	15.0997	.00438596	288	82944	16.9706	.00347222	348	121104	18.6548	.00287356
229	52441	15.1327	.00436681	289	83521	17.0000	.00346021	349	121801	18.6815	.00286533
230	52900	15.1658	.00434783	290	84100	17.0294	.00344828	350	122500	18.7083	.00285714
231	53361	15.1987	.00432900	291	84681	17.0587	.00343643	351	123201	18.7350	.00284900
232	53824	15.2315	.00431034	292	85264	17.0880	.00342466	352	123904	18.7617	.00284091
233	54289	15.2643	.00429185	293	85849	17.1172	.00341297	353	124609	18.7883	.00283286
234	54756	15.2971	.00427350	294	86436	17.1464	.00340136	354	125316	18.8149	.00282486
235	55225	15.3297	.00425532	295	87025	17.1756	.00338983	355	126025	18.8414	.00281690
236	55696	15.3623	.00423729	296	87616	17.2047	.00337838	356	126736	18.8680	.00280899
237	56169	15.3948	.00421941	297	88209	17.2337	.00336700	357	127449	18.8944	.00280112
238	56644	15.4272	.00420168	298	88804	17.2627	.00335570	358	128164	18.9209	.00279330
239	57121	15.4596	.00418410	299	89401	17.2916	.00334448	359	128881	18.9473	.00278552
240	57600	15.4919	.00416667	300	90000	17.3205	.00333333	360	129600	18.9737	.00277778

(continued)

Table Q. *(continued)*

N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N
361	130321	19.0000	.00277008	421	177241	20.5183	.00237530	481	231361	21.9317	.00207900
362	131044	19.0263	.00276243	422	178084	20.5426	.00236967	482	232324	21.9545	.00207469
363	131769	19.0526	.00275482	423	178929	20.5670	.00236407	483	233289	21.9773	.00207039
364	132496	19.0788	.00274725	424	179776	20.5913	.00235849	484	234256	22.0000	.00206612
365	133225	19.1050	.00273973	425	180625	20.6155	.00235294	485	235225	22.0227	.00206186
366	133956	19.1311	.00273224	426	181476	20.6398	.00234742	486	236196	22.0454	.00205761
367	134689	19.1572	.00272480	427	182329	20.6640	.00234192	487	237169	22.0681	.00205339
368	135424	19.1833	.00271739	428	183184	20.6882	.00233645	488	238144	22.0907	.00204918
369	136161	19.2094	.00271003	429	184041	20.7123	.00233100	489	239121	22.1133	.00204499
370	136900	19.2354	.00270270	430	184900	20.7364	.00232558	490	240100	22.1359	.00204082
371	137641	19.2614	.00269542	431	185761	20.7605	.00232019	491	241081	22.1585	.00203666
372	138384	19.2873	.00268817	432	186624	20.7846	.00231481	492	242064	22.1811	.00203252
373	139129	19.3132	.00268097	433	187489	20.8087	.00230947	493	243049	22.2036	.00202840
374	139876	19.3391	.00267380	434	188356	20.8327	.00230415	494	244036	22.2261	.00202429
375	140625	19.3649	.00266667	435	189225	20.8567	.00229885	495	245025	22.2486	.00202020
376	141376	19.3907	.00265957	436	190096	20.8806	.00229358	496	246016	22.2711	.00201613
377	142129	19.4165	.00265252	437	190969	20.9045	.00228833	497	247009	22.2935	.00201207
378	142884	19.4422	.00264550	438	191844	20.9284	.00228311	498	248004	22.3159	.00200803
379	143641	19.4679	.00263852	439	192721	20.9523	.00227790	499	249001	22.3383	.00200401
380	144400	19.4936	.00263158	440	193600	20.9762	.00227273	500	250000	22.3607	.00200000
381	145161	19.5192	.00262467	441	194481	21.0000	.00226757	501	251001	22.3830	.00199601
382	145924	19.5448	.00261780	442	195364	21.0238	.00226244	502	252004	22.4054	.00199203
383	146689	19.5704	.00261097	443	196249	21.0476	.00225734	503	253009	22.4277	.00198807
384	147456	19.5959	.00260417	444	197136	21.0713	.00225225	504	254016	22.4499	.00198413
385	148225	19.6214	.00259740	445	198025	21.0950	.00224719	505	255025	22.4722	.00198020
386	148996	19.6469	.00259067	446	198916	21.1187	.00224215	506	256036	22.4944	.00197628
387	149769	19.6723	.00258398	447	199809	21.1424	.00223714	507	257049	22.5167	.00197239
388	150544	19.6977	.00257732	448	200704	21.1660	.00223214	508	258064	22.5389	.00196850
389	151321	19.7231	.00257069	449	201601	21.1896	.00222717	509	259081	22.5610	.00196464
390	152100	19.7484	.00256410	450	202500	21.2132	.00222222	510	260100	22.5832	.00196078
391	152881	19.7737	.00255754	451	203401	21.2368	.00221729	511	261121	22.6053	.00195695
392	153664	19.7990	.00255102	452	204304	21.2603	.00221239	512	262144	22.6274	.00195312
393	154449	19.8242	.00254453	453	205209	21.2838	.00220751	513	263169	22.6495	.00194932
394	155236	19.8494	.00253807	454	206116	21.3073	.00220264	514	264196	22.6716	.00194553
395	156025	19.8746	.00253165	455	207025	21.3307	.00219870	515	265225	22.6936	.00194175
396	156816	19.8997	.00252525	456	207936	21.3542	.00219298	516	266256	22.7156	.00193798
397	157609	19.9249	.00251889	457	208849	21.3776	.00218818	517	267289	22.7376	.00193424
398	158404	19.9499	.00251256	458	209764	21.4009	.00218341	518	268324	22.7596	.00193050
399	159201	19.9750	.00250627	459	210681	21.4243	.00217865	519	269361	22.7816	.00192678
400	160000	20.0000	.00250000	460	211600	21.4476	.00217391	520	270400	22.8035	.00192308
401	160801	20.0250	.00249377	461	212521	21.4709	.00216920	521	271441	22.8254	.00191939
402	161604	20.0499	.00248756	462	213444	21.4942	.00216450	522	272484	22.8473	.00191571
403	162409	20.0749	.00248139	463	214369	21.5174	.00215983	523	273529	22.8692	.00191205
404	163216	20.0998	.00247525	464	215296	21.5407	.00215517	524	274576	22.8910	.00190840
405	164025	20.1246	.00246914	465	216225	21.5639	.00215054	525	275625	22.9129	.00190476
406	164836	20.1494	.00246305	466	217156	21.5870	.00214592	526	276676	22.9347	.00190114
407	165649	20.1742	.00245700	467	218089	21.6102	.00214133	527	277729	22.9565	.00189753
408	166464	20.1990	.00245098	468	219024	21.6333	.00213675	528	278784	22.9783	.00189394
409	167281	20.2237	.00244499	469	219961	21.6564	.00213220	529	279841	23.0000	.00189036
410	168100	20.2485	.00243902	470	220900	21.6795	.00212766	530	280900	23.0217	.00188679
411	168921	20.2731	.00243309	471	221841	21.7025	.00212314	531	281961	23.0434	.00188324
412	169744	20.2978	.00242718	472	222784	21.7256	.00211864	532	283024	23.0651	.00187970
413	170569	20.3224	.00242131	473	223729	21.7486	.00211416	533	284089	23.0868	.00187617
414	171396	20.3470	.00241546	474	224676	21.7715	.00210970	534	285156	23.1084	.00187266
415	172225	20.3715	.00240964	475	225625	21.7945	.00210526	535	286225	23.1301	.00186916
416	173056	20.3961	.00240385	476	226576	21.8174	.00210084	536	287296	23.1517	.00186567
417	173889	20.4206	.00239808	477	227529	21.8403	.00209644	537	288369	23.1733	.00186220
418	174724	20.4450	.00239234	478	228484	21.8632	.00209205	538	289444	23.1948	.00185874
419	175561	20.4695	.00238663	479	229441	21.8861	.00208768	539	290521	23.2164	.00185529
420	176400	20.4939	.00238095	480	230400	21.9089	.00208333	540	291600	23.2379	.00185185

N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N
541	292681	23.2594	.00184843	601	361201	24.5153	.00166389	661	436921	25.7099	.00151286
542	293764	23.2809	.00184502	602	302404	24.5357	.00166113	662	438244	25.7294	.00151057
543	294849	23.3024	.00184162	603	363609	24.5561	.00165837	663	439569	25.7488	.00150830
544	295936	23.3238	.00183824	604	364816	24.5764	.00165563	664	440896	25.7682	.00150602
545	297025	23.3452	.00183486	605	366025	24.5967	.00165289	665	442225	25.7876	.00150376
566	298116	23.3666	.00183150	606	367236	24.6171	.00165017	666	443556	25.8070	.00150150
547	299209	23.3880	.00182815	607	368449	24.6374	.00164745	667	444889	25.8263	.00149925
548	300304	23.4094	.00182482	608	369664	24.6577	.00164474	668	446224	25.8457	.00149701
549	301401	23.4307	.00182149	609	370881	24.6779	.00164204	669	447561	25.8650	.00149477
550	302500	23.4521	.00181818	610	372100	24.6982	.00163934	670	448900	25.8844	.00149254
551	303601	23.4734	.00181488	611	373321	24.7184	.00163666	671	450241	25.9037	.00149031
552	304704	23.4947	.00181159	612	374544	24.7386	.00163399	672	451584	25.9230	.00148810
553	305809	23.5160	.00180832	613	375769	24.7588	.00163132	673	452929	25.9422	.00148588
554	306916	23.5372	.00180505	614	376996	24.7790	.00162866	674	454276	25.9615	.00148368
555	308025	23.5584	.00180180	615	378225	24.7992	.00162602	675	455625	25.9808	.00148148
556	309136	23.5797	.00179856	616	379456	24.8193	.00162338	676	456976	26.0000	.00147929
557	310249	23.6008	.00179533	617	380689	24.8395	.00162075	677	458329	26.0192	.00147710
558	311364	23.6220	.00179211	618	381924	24.8596	.00161812	678	459684	26.0384	.00147493
559	312481	23.6432	.00178891	619	383161	24.8797	.00161551	679	461041	26.0576	.00147275
560	313600	23.6643	.00178571	620	384400	24.8998	.00161290	680	462400	26.0768	.00147059
561	314721	23.6854	.00178253	621	385641	24.9199	.00161031	681	463761	26.0960	.00146843
562	315844	23.7065	.00177936	622	386884	24.9399	.00160772	682	465124	26.1151	.00146628
563	316969	23.7276	.00177620	623	388129	24.9600	.00160514	683	466489	26.1343	.00146413
564	318096	23.7487	.00177305	624	389376	24.9800	.00160256	684	467856	26.1534	.00146199
565	319225	23.7697	.00176991	625	390625	25.0000	.00160000	685	469225	26.1725	.00145985
566	320356	23.7908	.00176678	626	391876	25.0200	.00159744	686	470596	26.1916	.00145773
567	321489	23.8118	.00176367	627	393129	25.0400	.00159490	687	471969	26.2107	.00145560
568	322624	23.8328	.00176056	628	394384	25.0599	.00159236	688	473344	26.2298	.00145349
569	323761	23.8537	.00175747	629	395641	25.0799	.00158983	689	474721	26.2488	.00145138
570	324900	23.8747	.00175439	630	396900	25.0998	.00158730	690	476100	26.2679	.00144928
571	326041	23.8956	.00175131	631	398161	25.1197	.00158479	691	477481	26.2869	.00144718
572	327184	23.9165	.00164825	632	399424	25.1396	.00158228	692	478864	26.3059	.00144509
573	328329	23.9374	.00174520	633	400689	25.1595	.00157978	693	480249	26.3249	.00144300
574	329476	23.9583	.00174216	634	401956	25.1794	.00157729	694	481636	26.3439	.00144092
575	330625	23.9792	.00173913	635	403225	25.1992	.00157480	695	483025	26.3629	.00143885
576	331776	24.0000	.00173611	636	404496	25.2190	.00157233	696	484416	26.3818	.00143678
577	332929	24.0208	.00173310	637	405769	25.2389	.00156986	697	485809	26.4008	.00143472
578	334084	24.0416	.00173010	638	407044	25.2587	.00156740	698	487204	26.4197	.00143266
579	335241	24.0624	.00172712	639	408321	25.2784	.00156495	699	488601	26.4386	.00143062
580	336400	24.0832	.00172414	640	409600	25.2982	.00156250	700	490000	26.4575	.00142857
581	337561	24.1039	.00172117	641	410881	25.3180	.00156006	701	491401	26.4764	.00142653
582	338724	24.1247	.00171821	642	412164	25.3377	.00155763	702	492804	26.4953	.00142450
583	339889	24.1454	.00171527	643	413449	25.3574	.00155521	703	494209	26.5141	.00142248
584	341056	24.1661	.00171233	644	414736	25.3772	.00155280	704	495616	26.5330	.00142045
585	342225	24.1868	.00170940	645	416025	25.3969	.00155039	705	497025	26.5518	.00141844
586	343396	24.2074	.00170648	646	417316	25.4165	.00154799	706	498436	26.5707	.00141643
587	344569	24.2281	.00170358	647	418609	25.4362	.00154560	707	499849	26.5895	.00141443
588	345744	24.2487	.00170068	648	419904	25.4558	.00154321	708	501264	26.6083	.00141243
589	346921	24.2693	.00169779	649	421201	25.4755	.00154083	709	502681	26.6271	.00141044
590	348100	24.2899	.00169492	650	422500	25.4951	.00153846	710	504100	26.6458	.00140845
591	349281	24.3105	.00169205	651	423801	25.5147	.00153610	711	505521	26.6646	.00140647
592	350464	24.3311	.00168919	652	425104	25.5343	.00153374	712	506944	26.6833	.00140449
593	351649	24.3516	.00168634	653	426409	25.5539	.00153139	713	508369	26.7021	.00140252
594	352836	24.3721	.00168350	654	427716	25.5734	.00152905	714	509796	26.7208	.00140056
595	354025	24.3926	.00168067	655	429025	25.5930	.00152672	715	511225	26.7395	.00139860
596	355216	24.4131	.00167785	656	430336	25.6125	.00152439	716	512656	26.7582	.00139665
597	356409	24.4336	.00167504	657	431649	25.6320	.00152207	717	514089	26.7769	.00139470
598	357604	24.4540	.00167224	658	432964	25.6515	.00151976	718	515524	26.7955	.00139276
599	358801	24.4745	.00166945	659	434281	25.6710	.00151745	719	516961	26.8142	.00139082
600	360000	24.4949	.00166667	660	435600	25.6905	.00151515	720	518400	26.8328	.00138889

(continued)

Table Q. *(continued)*

N	N²	√N	1/N	N	N²	√N	1/N	N	N²	√N	1/N
721	519841	26.8514	.00138696	781	609961	27.9464	.00128041	841	707281	29.0000	.00118906
722	521284	26.8701	.00138504	782	611524	27.9643	.00127877	842	708964	29.0172	.00118765
723	522729	26.8887	.00138313	783	613089	27.9821	.00127714	843	710649	29.0345	.00118624
724	524176	26.9072	.00138122	784	614656	28.0000	.00127551	844	712336	29.0517	.00118483
725	525625	26.9258	.00137931	785	616225	28.0179	.00127389	845	714025	29.0689	.00118343
726	527076	26.9444	.00137741	786	617796	28.0357	.00127226	846	715716	29.0861	.00118203
727	528529	26.9629	.00137552	787	619369	28.0535	.00127065	847	717409	29.1033	.00118064
728	529984	26.9815	.00137363	788	620944	28.0713	.00126904	848	719104	29.1204	.00117925
729	531441	27.0000	.00137174	789	622521	28.0891	.00126743	849	720801	29.1376	.00117786
730	532900	27.0185	.00136986	790	624100	28.1069	.00126582	850	722500	29.1548	.00117647
731	534361	27.0370	.00136799	791	625681	28.1247	.00126422	851	724201	29.1719	.00117509
732	535824	27.0555	.00136612	792	627264	28.1425	.00126263	852	725904	29.1890	.00117371
733	537289	27.0740	.00136426	793	628849	28.1603	.00126103	853	727609	29.2062	.00117233
734	538756	27.0924	.00136240	794	630436	28.1780	.00125945	854	729316	29.2233	.00117096
735	540225	27.1109	.00136054	795	632025	28.1957	.00125786	855	731025	29.2404	.00116959
736	541696	27.1293	.00135870	796	633616	28.2135	.00125628	856	732736	29.2575	.00116822
737	543169	27.1477	.00135685	797	635209	28.2312	.00125471	857	734449	29.2746	.00116686
738	544644	27.1662	.00135501	798	636804	28.2489	.00125313	858	736164	29.2916	.00116550
739	546121	27.1846	.00135318	799	638401	28.2666	.00125156	859	737881	29.3087	.00116414
740	547600	27.2029	.00135135	800	640000	28.2843	.00125000	860	739600	29.3258	.00116279
741	549081	27.2213	.00134953	801	641601	28.3019	.00124844	861	741321	29.3428	.00116144
742	550564	27.2397	.00134771	802	643204	28.3196	.00124688	862	743044	29.3598	.00116009
743	552049	27.2580	.00134590	803	644809	28.3373	.00124533	863	744769	29.3769	.00115875
744	553536	27.2764	.00134409	804	646416	28.3549	.00124378	864	746496	29.3939	.00115741
745	555025	27.2947	.00134228	805	648025	28.3725	.00124224	865	748225	29.4109	.00115607
746	556516	27.3130	.00134048	806	649636	28.3901	.00124069	866	749956	29.4279	.00115473
747	558009	27.3313	.00133869	807	651249	28.4077	.00123916	867	751689	29.4449	.00115340
748	559504	27.3496	.00133690	808	625864	28.4253	.00123762	868	753424	29.4618	.00115207
749	561001	27.3679	.00133511	809	654481	28.4429	.00123609	869	755161	29.4788	.00115075
750	562500	27.3861	.00133333	810	656100	28.4605	.00123457	870	756900	29.4958	.00114943
751	564001	27.4044	.00133156	811	657721	28.4781	.00123305	871	758641	29.5127	.00114811
752	565504	27.4226	.00132979	812	659344	28.4956	.00123153	872	760384	29.5296	.00114679
753	567009	27.4408	.00132802	813	660969	28.5132	.00123001	873	762129	29.5466	.00114548
754	568516	27.4591	.00132626	814	662596	28.5307	.00122850	874	763876	29.5635	.00114416
755	570025	27.4773	.00132450	815	664225	28.5482	.00122699	875	765625	29.5804	.00114286
756	571536	27.4955	.00132275	816	665856	28.5657	.00122549	876	767376	29.5973	.00114155
757	573049	27.5136	.00132100	817	667489	28.5832	.00122399	877	769129	29.6142	.00114025
758	574564	27.5318	.00131926	818	669124	28.6007	.00122249	878	770884	29.6311	.00113895
759	576081	27.5500	.00131752	819	670761	28.6182	.00122100	879	772641	29.6479	.00113766
760	577600	27.5681	.00131579	820	672400	28.6356	.00121951	880	774400	29.6848	.00113636
761	579121	27.5862	.00131406	821	674041	28.6531	.00121803	881	776161	29.6816	.00113507
762	580644	27.6043	.00131234	822	675684	28.6705	.00121655	882	777924	29.6985	.00113379
763	582169	27.6225	.00131062	823	677329	28.6880	.00121507	883	779689	29.7153	.00113250
764	583696	27.6405	.00130890	824	678976	28.7054	.00121359	884	781456	29.7321	.00113122
765	585225	27.6586	.00130719	825	680625	28.7228	.00121212	885	783225	29.7489	.00112994
766	586756	27.6767	.00130548	826	682276	28.7402	.00121065	886	784996	29.7658	.00112867
767	588289	27.6948	.00130378	827	683929	28.7576	.00120919	887	786769	29.7825	.00112740
768	589824	27.7128	.00130208	828	685584	28.7750	.00120773	888	788544	29.7993	.00112613
769	591361	27.7308	.00130039	829	687241	28.7924	.00120627	889	790321	29.8161	.00112486
770	592900	27.7489	.00129870	830	688900	28.8097	.00120482	890	792100	29.8329	.00112360
771	594441	27.7669	.00129702	831	690561	28.8271	.00120337	891	793881	29.8496	.00112233
772	595984	27.7849	.00129534	832	692224	28.8444	.00120192	892	795664	29.8664	.00112108
773	597529	27.8029	.00129366	833	693889	28.8617	.00120048	893	797449	29.8831	.00111982
774	599076	27.8209	.00129199	834	695556	28.8791	.00119904	894	799236	29.8998	.00111857
775	600625	27.8388	.00129032	835	697225	28.8964	.00119760	895	801025	29.9166	.00111732
776	602176	27.8568	.00128866	836	698896	28.9137	.00119617	896	802816	29.9333	.00111607
777	603729	27.8747	.00128700	837	700569	28.9310	.00119474	897	804609	29.9500	.00111483
778	605284	27.8927	.00128535	838	702244	28.9482	.00119332	898	806404	29.9666	.00111359
779	606841	27.9106	.00128370	839	703921	28.9655	.00119190	899	808201	29.9833	.00111235
780	608400	27.9285	.00128205	840	705600	28.9828	.00119048	900	810000	30.0000	.00111111

Table Q. *(concluded)*

N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N	N	N^2	$\sqrt{N}$	1/N
901	811801	30.0167	.00110988	936	876096	30.5941	.00106838	971	942841	31.1609	.00102987
902	813604	30.0333	.00110865	937	877969	30.6105	.00106724	972	944784	31.1769	.00102881
903	815409	30.0500	.00110742	938	879844	30.6268	.00106610	973	946729	31.1929	.00102775
904	817216	30.0666	.00110619	939	881721	30.6431	.00106496	974	948676	31.2090	.00102669
905	819025	30.0832	.00110497	940	883600	30.6594	.00106383	975	950625	31.2250	.00102564
906	820836	30.0998	.00110375	941	885481	30.6757	.00106270	976	952576	31.2410	.00102459
907	822649	30.1164	.00110254	942	887364	30.6920	.00106157	977	954529	31.2570	.00102354
908	824464	30.1330	.00110132	943	889249	30.7083	.00106045	978	956484	31.2730	.00102249
909	826281	30.1496	.00110011	944	891136	30.7246	.00105932	979	958441	31.2890	.00102145
910	828100	30.1662	.00109890	945	893025	30.7409	.00105820	980	960400	31.3050	.00102041
911	829921	30.1828	.00109769	946	894916	30.7571	.00105708	981	962361	31.3209	.00101937
912	831744	30.1993	.00109649	947	896809	30.7734	.00105597	982	964324	31.3369	.00101833
913	833569	30.2159	.00109529	948	898704	30.7896	.00105485	983	966289	31.3528	.00101729
914	835396	30.2324	.00109409	949	900601	30.8058	.00105374	984	968256	31.3688	.00101626
915	837225	30.2490	.00109290	950	902500	30.8221	.00105263	985	970225	31.3847	.00101523
916	839056	30.2655	.00109170	951	904401	30.8383	.00105152	986	972196	31.4006	.00101420
917	840889	30.2820	.00109051	952	906304	30.8545	.00105042	987	974169	31.4166	.00101317
918	842724	30.2985	.00108932	953	908209	30.8707	.00104932	988	976144	31.4325	.00101215
919	844561	30.3150	.00108814	954	910116	30.8869	.00104822	989	978121	31.4484	.00101112
920	846400	30.3315	.00108696	955	912025	30.9031	.00104712	990	980100	31.4643	.00101010
921	848241	30.3480	.00108578	956	913936	30.9192	.00104603	991	982081	31.4802	.00100908
922	850084	30.3645	.00108460	957	915849	30.9354	.00104493	992	984064	31.4960	.00100806
923	851929	30.3809	.00108342	958	917764	30.9516	.00104384	993	986049	31.5119	.00100705
924	853776	30.3974	.00108225	959	919681	30.9677	.00104275	994	988036	31.5278	.00100604
925	855625	30.4138	.00108108	960	921600	30.9839	.00104167	995	990025	31.5436	.00100503
926	857476	30.4302	.00107991	961	923521	31.0000	.00104058	996	992016	31.5595	.00100402
927	859329	30.4467	.00107875	962	925444	31.0161	.00103950	997	994009	31.5753	.00103842
928	861184	30.4631	.00107759	963	927369	31.0322	.00103842	998	996004	31.5911	.00100200
929	863041	30.4795	.00107643	964	929296	31.0483	.00103734	999	998001	31.6070	.00100100
930	864900	30.4959	.00107527	965	931225	31.0644	.00103627	1000	1000000	31.6228	.00100000
931	866761	30.5123	.00107411	966	933156	31.0805	.00103520				
932	868624	30.5287	.00107296	967	935089	31.0966	.00103413				
933	870489	30.5450	.00107181	968	937024	31.1127	.00103306				
934	872356	30.5614	.00107066	969	938961	31.1288	.00103199				
935	874225	30.5778	.00106952	970	940900	31.1448	.00103093				

Table R. Random Digits

Row number										
00000	10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
00001	37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
00002	08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
00003	99019	02529	09376	70715	38311	31165	88676	74397	04436	27659
00004	12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
00005	66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
00006	31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
00007	85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
00008	63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
00010	98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
00011	11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
00012	83452	99634	06288	98033	13746	70078	18475	40610	68711	77817
00013	88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
00015	65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
00016	80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
00017	74350	99817	77402	77214	43236	00210	45521	64237	96286	02655
00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
00019	09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
00020	91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
00021	80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
00023	12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
00024	63606	49329	16505	34484	40219	52563	43651	77082	07207	31790
00025	61196	90446	26457	47774	51924	33729	65394	59593	42582	60527
00026	15474	45266	95270	79953	59367	83848	82396	10118	33211	59466
00027	94557	28573	67897	54387	54622	44431	91190	42592	92927	45973
00028	42481	16213	97344	08721	16868	48767	03071	12059	25701	46670
00029	23523	78317	73208	89837	68935	91416	26252	29663	05522	82562
00030	04493	52494	75246	33824	45862	51025	61962	79335	65337	12472
00031	00549	97654	64051	88159	96119	63896	54692	82391	23287	29529
00032	35963	15307	26898	09354	33351	35462	77974	50024	90103	39333
00033	59808	08391	45427	26842	83609	49700	13021	24892	78565	20106
00034	46058	85236	01390	92286	77281	44077	93910	83647	70617	42941
00035	32179	00597	87379	25241	05567	07007	86743	17157	85394	11838
00036	69234	61406	20117	45204	15956	60000	18743	92423	97118	96338
00037	19565	41430	01758	75379	40419	21585	66674	36806	84962	85207
00038	45155	14938	19476	07246	43667	94543	59047	90033	20826	69541
00039	94864	31994	36168	10851	34888	81553	01540	35456	05014	51176
00040	98086	24826	45240	28404	44999	08896	39094	73407	35441	31880
00041	33185	16232	41941	50949	89435	48581	88695	41994	37548	73043
00042	80951	00406	96382	70774	20151	23387	25016	25298	94624	61171
00043	79752	49140	71961	28296	69861	02591	74852	20539	00387	59579
00044	18633	32537	98145	06571	31010	24674	05455	61427	77938	91936
00045	74029	43902	77557	32270	97790	17119	52527	58021	80814	51748
00046	54178	45611	80993	37143	05335	12969	56127	19255	36040	90324
00047	11664	49883	52079	84827	59381	71539	09973	33440	88461	23356
00048	48324	77928	31249	64710	02295	36870	32307	57546	15020	09994
00049	69074	94138	87637	91976	35584	04401	10518	21615	01848	76938
00050	09188	20097	32825	39527	04220	86304	83389	87374	64278	58044
00051	90045	85497	51981	50654	94938	81997	91870	76150	68476	64659
00052	73189	50207	47677	26269	62290	64464	27124	67018	41361	82760
00053	75768	76490	20971	87749	90429	12272	95375	05871	93823	43178
00054	54016	44056	66281	31003	00682	27398	20714	53295	07706	17813
00055	08358	69910	78542	42785	13661	58873	04618	97553	31223	08420
00056	28306	03264	81333	10591	40510	07893	32604	60475	94119	01840
00057	53840	86233	81594	13628	51215	90290	28466	68795	77762	20791
00058	91757	53741	61613	62669	50263	90212	55781	76514	83483	47055
00059	89415	92694	00397	58391	12607	17646	48949	72306	94541	37408

Table R. *(continued)*

Row number										
00060	77513	03820	86864	29901	68414	82774	51908	13980	72893	55507
00061	19502	37174	69979	20288	55210	29773	74287	75251	65344	67415
00062	21818	59313	93278	81757	05686	73156	07082	85046	31853	38452
00063	51474	66499	68107	23621	94049	91345	42836	09191	08007	45449
00064	99559	68331	62535	24170	69777	12830	74819	78142	43860	72834
00065	33713	48007	93584	72869	51926	64721	58303	29822	93174	93972
00066	85274	86893	11303	22970	28834	34137	73515	90400	71148	43643
00067	84133	89640	44035	52166	73852	70091	61222	60561	62327	18423
00068	56732	16234	17395	96131	10123	91622	85496	57560	81604	18880
00069	65138	56806	87648	85261	34313	65861	45875	21069	85644	47277
00070	38001	02176	81719	11711	71602	92937	74219	64049	65584	49698
00071	37402	96397	01304	77586	56271	10086	47324	62605	40030	37438
00072	97125	40348	87083	31417	21815	39250	75237	62047	15501	29578
00073	21826	41134	47143	34072	64638	85902	49139	06441	03856	54552
00074	73135	42742	95719	09035	85794	74296	08789	88156	64691	19202
00075	07638	77929	03061	18072	96207	44156	23821	99538	04713	66994
00076	60528	83441	07954	19814	59175	20695	05533	52139	61212	06455
00077	83596	35655	06958	92983	05128	09719	77433	53783	92301	50498
00078	10850	62746	99599	10507	13499	06319	53075	71839	06410	19362
00079	39820	98952	43622	63147	64421	80814	43800	09351	31024	73167
00080	59580	06478	75569	78800	88835	54486	23768	06156	04111	08408
00081	38508	07341	23793	48763	90822	97022	17719	04207	95954	49953
00082	30692	70668	94688	16127	56196	80091	82067	63400	05462	69200
00083	65443	95659	18238	27437	49632	24041	08337	65676	96299	90836
00084	27267	50264	13192	72294	07477	44606	17985	48911	97341	30358
00085	91307	06991	19072	24210	36699	53728	28825	35793	28976	66252
00086	68434	94688	84473	13622	62126	98408	12843	82590	09815	93146
00087	48908	15877	54745	24591	35700	04754	83824	52692	54130	55160
00088	06913	45197	42672	78601	11883	09528	63011	98901	14974	40344
00089	10455	16019	14210	33712	91342	37821	88325	80851	43667	70883
00090	12883	97343	65027	61184	04285	01392	17974	15077	90712	26769
00091	21778	30976	38807	36961	31649	42096	63281	02023	08816	47449
00092	19523	59515	65122	59659	86283	68258	69572	13798	16435	91529
00093	67245	52670	35583	16563	79246	86686	76463	34222	26655	90802
00094	60584	47377	07500	37992	45134	26529	26760	83637	41326	44344
00095	53853	41377	36066	94850	58838	73859	49364	73331	96240	43642
00096	24637	38736	74384	89342	52623	07992	12369	18601	03742	83873
00097	83080	12451	38992	22815	07759	51777	97377	27585	51972	37867
00098	16444	24334	36151	99073	27493	70939	85130	32552	54846	54759
00099	60790	18157	57178	65762	11161	78576	45819	52979	65130	04860
00100	03991	10461	93716	16894	66083	24653	84609	58232	88618	19161
00101	38555	95554	32886	59780	08355	60860	29735	47762	71299	23853
00102	17546	73704	92052	46215	55121	29281	59076	07936	27954	58909
00103	32643	52861	95819	06831	00911	98936	76355	93779	80863	00514
00104	69572	68777	39510	35905	14060	40619	29549	69616	33564	60780
00105	24122	66591	27699	06494	14845	46672	61958	77100	90899	75754
00106	61196	30231	92962	61773	41839	55382	17267	70943	78038	70267
00107	30532	21704	10274	12202	39685	23309	10061	68829	55986	66485
00108	03788	97599	75867	20717	74416	53166	35208	33374	87539	08823
00109	48228	63379	85783	47619	53152	67433	35663	52972	16818	60311
00110	60365	94653	35075	33949	42614	29297	01918	28316	98953	73231
00111	83799	42402	56623	34442	34994	41374	70071	14736	09958	18065
00112	32960	07405	36409	83232	99385	41600	11133	07586	15917	06253
00113	19322	53845	57620	52606	66497	68646	78138	66559	19640	99413
00114	11220	94747	07399	37408	48509	23929	27482	45476	85244	35159
00115	31751	57260	68980	05339	15470	48355	88651	22596	03152	19121
00116	88492	99382	14454	04504	20094	98977	74843	93413	22109	78508
00117	30934	47744	07481	83828	73788	06533	28597	20405	94205	20380
00118	22888	48893	27499	98748	60530	45128	74022	84617	82037	10268
00119	78212	16993	35902	91386	44372	15486	65741	14014	87481	37220

(continued)

Row number										
00120	41849	84547	46850	52326	34677	58300	74910	64345	19325	81549
00121	46352	33049	69248	93460	45305	07521	61318	31855	14413	70951
00122	11087	96294	14013	31792	59747	67277	76503	34513	39663	77544
00123	52701	08337	56303	87315	16520	69676	11654	99893	02181	68161
00124	57275	36898	81304	48585	68652	27376	92852	55866	88448	03584
00125	20857	73156	70284	24326	79375	95220	01159	63267	10622	48391
00126	15633	84924	90415	93614	33521	26665	55823	47641	86225	31704
00127	92694	48297	39904	02115	59589	49067	66821	41575	49767	04037
00128	77613	19019	88152	00080	20554	91409	96277	48257	50816	97616
00129	38688	32486	45134	63545	59404	72059	43947	51680	43852	59693
00130	25163	01889	70014	15021	41290	67312	71857	15957	68971	11403
00131	65251	07629	37239	33295	05870	01119	92784	26340	18477	65622
00132	36815	43625	18637	37509	82444	99005	04921	73701	14707	93997
00133	64397	11692	05327	82162	20247	81759	45197	25332	83745	22567
00134	04515	25624	95096	67946	48460	85558	15191	18782	16930	33361
00135	83761	60873	43253	84145	60833	25983	01291	41349	20368	07126
00136	14387	06345	80854	09279	43529	06318	38384	74761	41196	37480
00137	51321	92246	80088	77074	88722	56736	66164	49431	66919	31678
00138	72472	00008	80890	18002	94813	31900	54155	83436	35352	54131
00139	05466	55306	93128	18464	74457	90561	72848	11834	79982	68416
00140	39528	72484	82474	25593	48545	35247	18619	13674	18611	19241
00141	81616	18711	53342	44276	75122	11724	74627	73707	58319	15997
00142	07586	16120	82641	22820	92904	13141	32392	19763	61199	67940
00143	90767	04235	13574	17200	69902	63742	78464	22501	18627	90872
00144	40188	28193	29593	88627	94972	11598	62095	36787	00441	58997
00145	34414	82157	86887	55087	19152	00023	12302	80783	32624	68691
00146	63439	75363	44989	16822	36024	00867	76378	41605	65961	73488
00147	67049	09070	93399	45547	94458	74284	05041	49807	20288	34060
00148	79495	04146	52162	90286	54158	34243	46978	35482	59362	95938
00149	91704	30552	04737	21031	75051	93029	47665	64382	99782	93478
00150	94015	46874	32444	48277	59820	96163	64654	25843	41145	42820
00151	74108	88222	88570	74015	25704	91035	01755	14750	48968	38603
00152	62880	87873	95160	59221	22304	90314	72877	17334	39283	04149
00153	11748	12102	80580	41867	17710	59621	06554	07850	73950	79552
00154	17944	05600	60478	03343	25852	58905	57216	39618	49856	99326
00155	66067	42792	95043	52680	46780	56487	09971	59481	37006	22186
00156	54244	91030	45547	70818	59849	96169	61459	21647	87417	17198
00157	30945	57589	31732	57260	47670	07654	46376	25366	94746	49580
00158	69170	37403	86995	90307	94304	71803	26825	05511	12459	91314
00159	08345	88975	35841	85771	08105	59987	87112	21476	14713	71181
00160	27767	43584	85301	88977	29490	69714	73035	41207	74699	09310
00161	13025	14338	54066	15243	47724	66733	47431	43905	31048	56699
00162	80217	36292	98525	24335	24432	24896	43277	58874	11466	16082
00163	10875	62004	90391	61105	57411	06368	53856	30743	08670	84741
00164	54127	57326	26629	19087	24472	88779	30540	27886	61732	75454
00165	60311	42824	37301	42678	45990	43242	17374	52003	70707	70214
00166	49739	71484	92003	98086	76668	73209	59202	11973	02902	33250
00167	78626	51594	16453	94614	39014	97066	83012	09832	25571	77628
00168	66692	13986	99837	00582	81232	44987	09504	96412	90193	79568
00169	44071	28091	07362	97703	76447	42537	98524	97831	65704	09514
00170	41468	85149	49554	17994	14924	39650	95294	00556	70481	06905
00171	94559	37559	49678	53119	70312	05682	66986	34099	74474	20740
00172	41615	70360	64114	58660	90850	64618	80620	51790	11436	38072
00173	50273	93113	41794	86861	24781	89683	55411	85667	77535	99892
00174	41396	80504	90670	08289	40902	05069	95083	06783	28102	57816
00175	25807	24260	71529	78920	72682	07385	90726	57166	98884	08583
00176	06170	97965	88302	98041	21443	41808	68984	83620	89747	98882
00177	60808	54444	74412	81105	01176	28838	36421	16489	18059	51061
00178	80940	44893	10408	36222	80582	71944	92638	40333	67054	16067
00179	19516	90120	46759	71643	13177	55292	21036	82808	77501	97427

Table R. *(concluded)*

Row number										
00180	49386	54480	23604	23554	21785	41101	91178	10174	29420	90438
00181	06312	88940	15995	69321	47458	64809	98189	81851	29651	84215
00182	60942	00307	11897	92674	40405	68032	96717	54244	10701	41393
00183	92329	98932	78284	46347	71209	92061	39448	93136	25722	08564
00184	77936	63574	31384	51924	85561	29671	58137	17820	22751	36518
00185	38101	77756	11657	13897	95889	57067	47648	13885	70669	93406
00186	39641	69457	91339	22502	92613	89719	11947	56203	19324	20504
00187	84054	40455	99396	63680	67667	60631	69181	96845	38525	11600
00188	47468	03577	57649	63266	24700	71594	14004	23153	69249	05747
00189	43321	31370	28977	23896	76479	68562	62342	07589	08899	05985
00190	64281	61826	18555	64937	13173	33365	78851	16499	87064	13075
00191	66847	70495	32350	02985	86716	38746	26313	77463	55387	72681
00192	72461	33230	21529	53424	92581	02262	78438	66276	18396	73538
00193	21032	91050	13058	16218	12470	56500	15292	76139	59526	52113
00194	95362	67011	06651	16136	01016	00857	55018	56374	35824	71708
00195	49712	97380	10404	55452	34030	60726	75211	10271	36633	68424
00196	58275	61764	97586	54716	50259	46345	87195	46092	26787	60939
00197	89514	11788	68224	23417	73959	76145	30342	40277	11049	72049
00198	15472	50669	48139	36732	46874	37088	63465	09819	58869	35220
00199	12120	86124	51247	44302	60883	52109	21437	36786	49226	77837

Answers to Selected Exercises

Answers to Selected Exercises

Chapter 1

1. a) Statistic b) Statistic c) Data
   d) Data e) Inference from data f) Inference from data
   g) Statistic

Chapter 2

1. a) $X = b - c - a$ b) $X = \dfrac{c - a}{2}$ c) $X = c - \dfrac{bY}{a}$
   d) $X = ac/b$ e) $X = acY/b$ f) $X = (C^2Y/b) - a$
   g) $X = a/c - b$ h) $X = \sqrt[9]{Y}$ i) $X = \dfrac{1}{\sqrt[5]{a}}$
2. a) $X = 3$ b) $X = 24$ c) $X = \frac{125}{7776}$ d) $X = 27$ e) $X = 7\frac{1}{9}$
3. a) 100.00 b) 46.41 c) 2.96 d) 0.01
   e) 16.46 f) 1.05 g) 86.21 h) 10.00
4. a) 0.03052 b) 0.084680 c) 0.12088
   d) 0.38 e) 0.111 f) 0.85714
5. b) 0.2308; 23.08% c) 0.7576; 75.76% d) 0.4706; 47.06%
6. a) 25 b) 60 c) 37
   d) 31 e) 60 f) 44
7. a) $\sum_{i=1}^{3} X_i$ b) $\sum_{i=1}^{N} X_i$ c) $\sum_{i=3}^{6} X_i^2$ d) $\sum_{i=4}^{N} X_i^2$
8. a) Ratio b) Ratio c) Nominal d) Ordinal e) Nominal
9. a) Ratio; continuous b) Interval; continuous
   c) Nominal; discontinuous d) Ordinal; discontinuous
10. a) 12.65 b) 4.00 c) 1.26 d) 0.40 e) 0.13
11. a) -0.5 to $+0.5$ b) 0.45 to 0.55 c) 0.95 to 1.05
    d) 0.485 to 0.495 e) -4.5 to -5.5 f) -4.45 to -4.55

15. a) Discrete b) Discrete c) Continuous
d) Continuous e) Continuous f) Discrete

16.

	a)	b)	c)
Business administration	20.00%	38.10	13.33
Education	75.00%	4.76	20.00
Humanities	57.14%	14.29	26.67
Science	28.57%	23.81	13.33
Social science	50.00%	19.05	26.67

d) 58.33% Males
41.67% Females

19. a) Continuous b) Discrete c) Continuous
d) Discrete e) Discrete

20. a)

Year	% Males	% Females
1940	51.33	48.67
1945	51.35	48.65
1950	51.31	48.69
1955	51.24	48.76
1960	51.20	48.80

b) Time ratios, fixed base (1940)

Year	Males	Females
1940	100.00	100.00
1945	115.92	115.84
1950	150.50	150.65
1955	171.12	171.80
1960	179.87	180.85

Chapter 3

1.

	True limits	Midpoint	Width
a)	7.5–12.5	10	5
b)	5.5–7.5	6.5	2
c)	(−0.5)–(+2.5)	1	3
d)	4.5–14.5	9.5	10
e)	(−1.5)–(−8.5)	−5	7
f)	2.45–3.55	3	1.1
g)	1.495–1.755	1.625	0.26
h)	(−3.5)–(+3.5)	0	7

3.

Class intervals	True limits	Midpoint	f	Cum f	Cum %
95–99	94.5–99.5	97	1	40	100.0%
90–94	89.5–94.5	92	3	39	97.5
85–89	84.5–89.5	87	4	36	90.0
80–84	79.5–84.5	82	8	32	80.0
75–79	74.5–79.5	77	11	24	60.0
70–74	69.5–74.5	72	4	13	32.5
65–69	64.5–69.5	67	3	9	22.5
60–64	59.5–64.5	62	3	6	15.0

3.

Class intervals	True limits	Midpoint	f	Cum f	Cum %
55–59	54.5–59.5	57	0	3	7.5
50–54	49.5–54.5	52	1	3	7.5
45–49	44.5–49.5	47	1	2	5.0
40–44	39.5–44.5	42	1	1	2.5

4. b) $i = 3$

Class intervals	f
96–98	1
93–95	1
90–92	2
87–89	3
84–86	2
81–83	7
78–80	4
75–77	7
72–74	1
69–71	3

Class intervals	f
66–68	3
63–65	1
60–62	2
57–59	0
54–56	0
51–53	1
48–50	0
45–47	1
42–44	0
39–41	1

c) $i = 10$

Class intervals	f
90–99	4
80–89	12
70–79	15
60–69	6
50–59	1
40–49	2

d) $i = 20$

Class intervals	f
80–99	16
60–79	21
40–59	3

6. a) $i = 7$

b) 14.5–21.5
7.5–14.5
0.5– 7.5

c) 18
11
4

8.

Class intervals	f
65–69	1
60–64	2
55–59	3
50–54	4
45–49	5
40–44	6
35–39	8

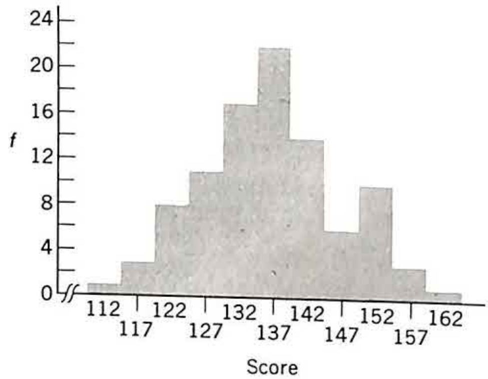
Class intervals	f
30–34	6
25–29	5
20–24	4
15–19	3
10–14	2
5–9	1

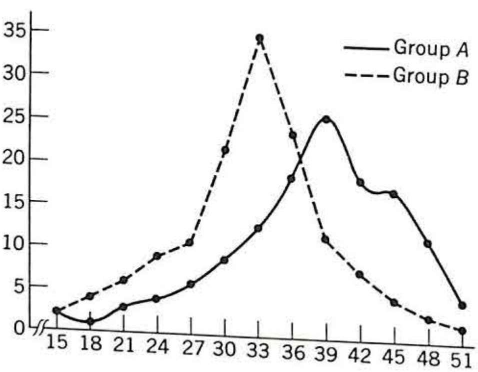
9.

Class intervals	f
63–67	3
58–62	0
53–57	7
48–52	0
43–47	11
38–42	0
33–37	14

Class intervals	f
28–32	0
23–27	9
18–22	0
13–17	5
8–12	0
3–7	1

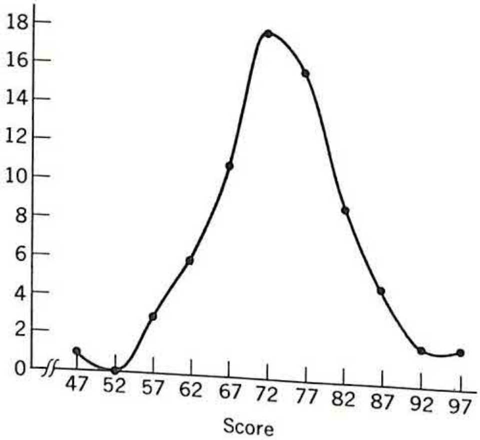
13.



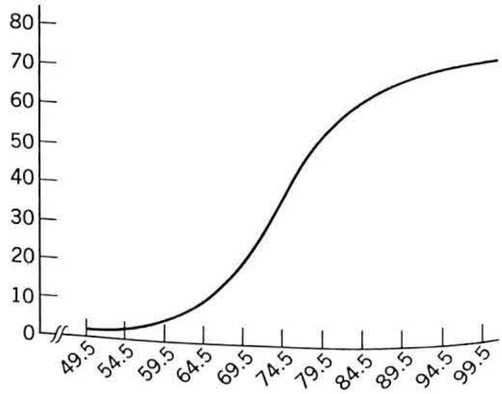


19. a) Positively skewed
b) Normal
c) Normal
d) Bimodal

20.



21.

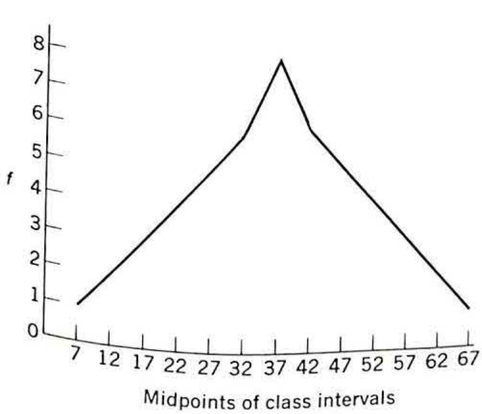


24. a) 34.62%

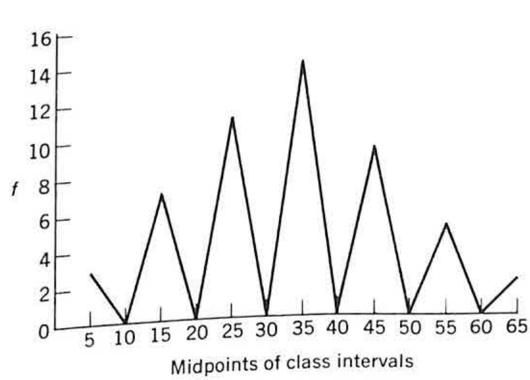
b) Class intervals	f	c) cf
92–94	1	52
89–91	0	51
86–88	2	51
83–85	0	49
80–82	0	49
77–79	3	49
74–76	5	46
71–73	7	41
68–70	5	34
65–67	10	29
62–64	11	19
59–61	4	8
56–58	4	4

d) Class intervals	f	cf
90–94	1	52
85–89	2	51
80–84	0	49
75–79	7	49
70–74	9	42
65–69	14	33
60–64	15	19
55–59	4	4

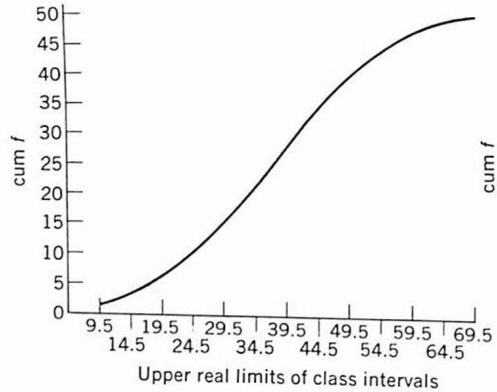
25.



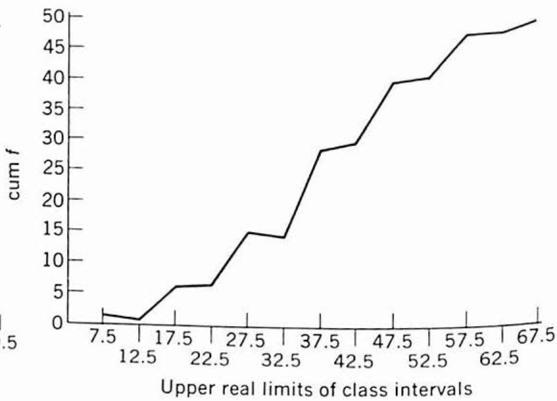
26.



27.



28.



Chapter 4

2. a) 26.36 b) 46.82 c) 88.45
4. a) 103.88 b) 113.03 c) 124.78
6. a) 81.02 b) 94.76 c) 88.04
7. a) $\frac{16.9}{110} = 15.36\%$ b) $\frac{93.1}{110} = 84.64\%$ c) 2.45%
   d) 0.18% e) 26.73% f) 3.73%
8. b) 57, 76 c) Law school freshmen, 60
9. a) 2 b) 9.6 c) 19.2 d) 36 e) 42 f) 84.8
10. a) 2 b) 12 c) 17.4 d) 30 e) 41.2 f) 80
13. a) 17.8 b) 24.5 c) 37 d) 40.3 e) 47
14. a) 16.5 b) 24.7 c) 36.1 d) 43 e) 46.4

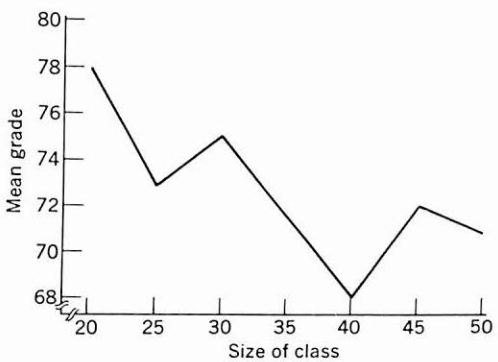
Chapter 5

1. a) $\overline{X} = 4.7$ Median = 4.5 Mode = 8
   b) $\overline{X} = 5.0$ Median = 5.0 Mode = 5.0
   c) $\overline{X} = 17.5$ Median = 4 Mode = 4
2. (c)
4. All measures of central tendency will be reduced by 5.
5. All measures of central tendency will be divided by 16.
6. Chapter 3, Problem 13
   $\overline{X} = 136.64$ Median = 136.32 Mode = 137
   Chapter 3, Problem 20
   $\overline{X} = 74.05$ Median = 73.81 Mode = 72

7. $\overline{X} = 5.0$; median unchanged, mode unchanged
8. a) Negative skew b) Positive skew
   c) No evidence of skew d) No evidence of skew
9. 7(c) Symmetrical 7(d) Bimodal
10. $\overline{X} = 5.0$: a) $\overline{X} = 7.0$ b) $\overline{X} = 3.0$ c) No change d) $\overline{X} = 10.0$ e) $\overline{X} = 2.5$
11. Group A: median; Group B: mean
12. b)

	Means f	Medians f
6	1	3
5	6	5
4	11	9
3	7	6
2	4	3
1	1	4

   d) Drawing from means: 1) $\frac{1}{30}$ 2) $\frac{1}{30}$ 3) $\frac{2}{30}$
   Drawing from medians: 1) $\frac{3}{30}$ 2) $\frac{4}{30}$ 3) $\frac{7}{30}$
13. $\overline{X} = 6.46$, Median $= 6.1$
14. a) \$3.03 b) \$3.12
15. a) 71.9 b)



17. $\overline{X} = -1.44$, Median $= -0.25$, Mode $= -5.00$
22. Raise median: (c)
    Raise mean: (b)

Chapter 6

1. $s^2 = 1.67$, $s = 1.29$
   a) No change b) No change c) Increase d) Increase e) Decrease
5. All scores have the same value.
6. a) 3.52 b) 2.31 c) 5.89 d) 0
8. a) 11 b) 9 c) 21 d) 1
10. Problem 8, Chapter 3: $\overline{X} = 37.00$, $s = 14.00$
    Problem 9, Chapter 3: $\overline{X} = 37.40$, $s = 14.22$

11. a) $\overline{X} = 63.10$ b) Range = 36, Interquartile range = 8.00, M.D. = 5.50
    c) $s = 7.89$, $s^2 = 62.19$

12. $\overline{X} = 8.80$, $s = 13.05$

13. $\overline{X} \doteq 68.25$, $s = 7.61$

16.

		January, 1965	January, 1966	May, 1965	May, 1966
a)	$\overline{X}$	35.58	38.65	76.90	70.42
b)	Range	40	44	41	34
	M.D.	7.78	7.33	7.47	7.81
c)	s	9.72	8.13	9.55	9.23
	s^2	94.55	66.05	91.23	85.18

Chapter 7

1. a) 0.94 b) −0.40 c) 0.00
   d) −1.32 e) 2.23 f) −2.53

2. a) 0.4798 b) 0.4713 c) 0.0987 d) 0.1554
   e) 0.4505 f) 0.4750 g) 0.4901 h) 0.4951

3. a) 1) 0.3413; 341 2) 0.4772; 477 3) 0.1915; 192 4) 0.4938; 494
   b) 1) 0.1587; 159 2) 0.0228; 23 3) 0.6915; 692 4) 0.9938; 994 5) 0.5000; 500
   c) 1) 0.1359; 136 2) 0.8351; 835 3) 0.6687; 669 4) 0.3023; 302

5. a) 40.13; 57.93 b) 60.26

6. (a),(b),(c),(d)

9. Test I, Test I

13. a)

	National League	American League
$\overline{X}$	43.08	39.75
s	7.81	8.71

c) American League, $V = 21.91\%$

Chapter 8

1. $r = 0.8466$

2. $r = 0.9107$

3. $r_{\text{rho}} = 0.906$

6. $r = 0.8260$

7. $r = 0$; $r_{\text{rho}} = 0.50$

11. $r = -0.7692$

12. $r = 0.7614$

13. a) $r = -0.8924$ b) $r = -0.8720$ c) $r = 0.9729$

14. a) $r_{\text{rho}} = -0.7947$ b) $r_{\text{rho}} = -0.6177$ c) $r_{\text{rho}} = 0.9557$

15. a) $r = 0.1068$ b) $r = 0.0734$
    c) Jan., 1965, $r_{\text{rho}} = -0.1700$
    Jan., 1966, $r_{\text{rho}} = -0.7253$
    May, 1965, $r_{\text{rho}} = 0.1898$
    May, 1966, $r_{\text{rho}} = 0.9329$
16. a) $r_{\text{rho}} = 0.1366$ b) $r_{\text{rho}} = 0.0814$
17. a) $r = 0.1412$
18. a) $r = 0.2987$ b) National League, $r_{\text{rho}} = 0.4172$
    American League, $r_{\text{rho}} = 0.5624$

Chapter 9

1. $Y' = 3 - 0.9\,(X - 3)$
2. a) 1.72 b) 118.48 c) $s_{\text{est}\,y} = 0.3928$
3. a) 1.59 b) $s_{\text{est}\,y} = 0.47$; $s_{\text{est}\,x} = 11.20$ c) 0.1296
4. As r increases the angle formed by the regression lines decreases. In the limiting cases ($r = \pm 1.00$), the regression lines are superimposed upon each other. When $r = 0$, the regression lines are at right angles to each other.
6. a) 0 b) 0.60 c) 1.20
   d) 1.5 e) -0.75 f) -1.20
7. a) $Y' = 89.17$
   b) The proportion (p) of area corresponding to a score of 110 or higher is 0.0031
   c) $X = 31.54$ d) $p = 0.4129$ e) $p < 0.00003$
   f) $X = 68.46$ g) 500, 93, 93, 907
8. a) $r = 0.9398$ b) 50, 40, 60, 70
10. a) 34.41 b) $p = 0.1949$ c) 48.64 d) $p = 0.9082$
    e) 57.11 f) $p = 0.4207$ g) $p = 0.3897$
11. a) $r = -0.8450$ b) (1) 3.78 (2) $p = 0.0084$

Review Section I

1. B. (1) 11.70 (2) 142.5 (3) 3.74
   C. (1) $r_{\text{rho}} = -0.106$ (2) $r = -0.185$
2. C. $\overline{X} = 7.961$, $s = 0.409$ D. $r = -0.7855$ 1) 39.273

Chapter 10

1. a) 0.1250 b) 0.1250 c) 0.3750 d) 0.5000
2. a) 0.0192 b) 0.0769 c) 0.3077 d) 0.4231
3. Problem 1: a) 7 to 1 against, b) 7 to 1 against
   Problem 2: a) 51 to 1 against, b) 12 to 1 against
4. a) 0.1667 b) 0.1667 c) 0.2778 d) 0.5000

5. a) 0.0046 b) 0.0278 c) 0.0139 d) 0.0556 e) 0.5787
6. a) 0.0129 b) 0.4135 c) 0.0100
7. a) 0.0156 b) 0.4219 c) 0.0123
8. a) 0.0838 b) 126.40 c) <68.64 and >131.36
   d) 0.4648 e) i) 0.0070 ii) 0.2160 iii) 0.0390
9. (b)
10. $p = \frac{1}{24}$
11. a) $p = 0.20$ b) $p = 0.04$ c) $p = 0.80$ d) $p = 0.64$
12. a) 0.6561 b) 0.0001 c) 0.3438
13. a) $p = 0.0456$ b) $p = 0.0021$ c) $p = 0.9544$
    d) $p = 0.9109$ e) $p = 0.0228$ f) $p = 0.0520$
14. a) $\frac{1}{3}$ b) $\frac{1}{2}$ c) $\frac{1}{6}$ d) 0
15. a) $\frac{1}{11}$ b) $\frac{1}{66}$ c) $\frac{14}{99}$
16. a) $\frac{1}{8}$ b) $\frac{1}{36}$ c) $\frac{16}{81}$
17. a) 40% b) 12% c) 16% d) 3.36%
18. a) $p = 0.3085$ b) $p = 0.2266$ c) $p = 0.4649$
    d) $p = 0.3104$ e) $p = 0.0947$ f) $p = 0.0947$

Chapter 11

6. a) $\frac{1}{1024}$ or 0.00098 b) 0.055 c) 17.18 to 1 against passing
7. a) H_0 b) H_1 c) H_1 d) H_0
8. a) Type II b) Type I c) No error d) No error
9. a) $p = 0.062$ b) $p = 0.773$ c) $p = 0.124$ d) $p = 0.938$
10. a) Type II b) Type II c) no error

Chapter 12

3. a) 21.71–26.29 b) 20.82–27.18
4. a) 23.28–24.72 b) 23.05–24.95
5. Reject H_0; $z = 2.68$ in which $z_{0.01} = \pm 2.58$
6. Accept H_0; $t = 2.618$ in which $t_{0.01} = \pm 2.831$, df $= 21$
8. Reject H_0; $t = 2.134$ in which $t_{0.05} = 1.833$ (one-tailed test), df $= 9$
9. a) $z = 2.73$, $p = 0.9968$ b) $z = -0.91$, $p = 0.1814$
   c) $z = 1.82$, $p = 0.0344$ d) $p = 0.6372$
10. a) 36.81–43.19 b) 35.79–44.21
11. a) $z = 1.66$ b) $z = 7.64$
12. $t = 2.083$, df $= 25$

13. 27.14 — 32.08

14. $t = 1.75$, df $= 625$

15. 263.34 — 273.66

16. a) $\mu = 6.0$, $\sigma = 3.42$

17. a) $p = \frac{1}{15}$ b) $p = \frac{1}{15}$ c) $p = \frac{8}{15}$ d) $p = 0$

18. a) $p = \frac{1}{12}$ b) $p = \frac{1}{12}$ c) $p = \frac{5}{9}$ d) $p = \frac{1}{36}$

19. $t = 1.826$, df $= 7$

20. $t = 2.379$, df $= 9$

23. a) $z = 1.11$ b) $z = 1.99$

25. No

26. Yes

Chapter 13

1. $t = 2.795$, reject H_0

2. $t = 1.074$, accept H_0

3. $t = 0.802$, accept H_0

4. a) $t = 3.522$, reject H_0 b) $F = 3.00$, accept H_0

5. a) $z = 1.41$, $p = 0.0793$ b) $z = -2.12$, $p = 0.9830$
   c) $z = -2.12$, $p = 0.0170$ d) $z = -5.66$, $p < 0.00003$

6. a) $z = -0.71$, $p = 0.7611$ b) $z = -1.06$, $p = 0.8554$
   c) $z = -1.41$, $p = 0.9207$ d) $z = -1.77$, $p = 0.9616$

8. $t = 2.148$, df $= 14$

9. $t = 2.913$, df $= 18$

10. $t = 2.304$, df $= 26$

11. $t = 1.361$, df $= 25$

12. $t = 1.275$, df $= 162$

13. $t = 2.224$, df $= 28$

14. $t = 0.583$, df $= 34$

15. $t = 0.643$, df $= 20$

16. a) False b) True c) False

18. a) $p = 0.7778$ b) $p = 0.6111$ c) $p = 0.3889$ d) $p = 0.2222$
    e) $p = 0.0833$ f) $p = 0$ g) $p = 0.4722$

19. $\mu_1 = 5.0$, $\sigma_1 = 2.24$
    $\mu_2 = 4.0$, $\sigma_2 = 2.24$

20. a) $z = -1.59$, $p = 0.9441$ b) $z = 0.00$, $p = 0.5000$
    c) $z = -1.59$, $p = 0.0559$ d) $z = -3.18$, $p = 0.0007$
    e) $z = 4.76$, $p = 0.00003$ f) $p = 0.5007$
    g) $p = 0.0559$

21. a) $z = -0.63$, $p = 0.2643$ b) $z = -0.94$, $p = 0.1736$
c) $z = -1.27$, $p = 0.1020$ d) $z = -1.92$, $p = 0.0274$

Chapter 14

1. $t = 3.027$, df $= 16$; reject H_0.
2. $t = 2.08$, df $= 32$; accept H_0.
3. $t = 1.571$, $A = 0.465$, df $= 9$; accept H_0.
4. r_{rho} between final standing and homeruns in American League is 0.04; for National League it is 0.36. Both correlations are too low to justify use of correlated samples.
7. a) $A = 0.048$ b) $A = 0.048$ c) $A = 0.778$
8. $A = 0.199$
10. Manufacturer A: $A = 0.195$
Manufacturer B: $A = 0.573$
11. $A = 0.216$
12. $A = 5.480$
13. $A = 0.210$
14. $A = 0.230$
15. $A = 0.422$

Chapter 15

1. $F = 36.99$; reject H_0.
2. All comparisons are significant at 0.01 level. It is clear that death rates are lowest in summer, next to lowest in spring, next to highest in fall, and highest in winter.
3. $F = 0.643$
5. $F = 0.30$, df $= 3/12$
6. $F = 10.43$, df $= 2/12$ HSD $= 11.89$
7. $F = 12.99$, df $= 5/18$ HSD $= 6.02$

REVIEW OF SECTION II INFERENTIAL STATISTICS

A. Parametric Tests

1. a) $\mu = 4$, $\sigma = 1.15$

b)

$\overline{X}$	f
2	1
2.5	4
3	10
3.5	16
4	19
4.5	16
5	10
5.5	4
6	1

c)

$\overline{X}$	p
2	0.0123
2.5	0.0494
3	0.1235
3.5	0.1975
4	0.2346
4.5	0.1975
5	0.1235
5.5	0.0494
6	0.0123

d) 0.82 e) 0.82
f) i) 0.0123; ii) 0.1852; iii) 0.3704
2. a) $t = 1.856$, df $= 8$; accept H_0. b) $t = 2.727$, df $= 18$; reject H_0.
3. a) $A = 0.210$, df $= 4$; reject H_0. b) $A = 0.111$, df $= 9$; reject H_0.
4. $F = 16.40$, df $= 2/27$

Chapter 16

1. $N_a = 30$
3. 28.77%
4. a) 1) 0; 2) $\beta = 0.1685$; 3) Power $= 0.8315$
   b) 1) α; 2) $\beta = 0$; 3) The concept of power does not apply.
5. a) 5.26% b) 10.20% c) 17.88% d) 28.77%

Chapter 17

1. a) $p = 0.070$ b) $p = 0.020$ c) $p = 0.040$
2. a) $p < 0.0003$ b) $p = 0.0003$ c) $p = 0.1335$
4. Inflated N
5. $\chi^2 = 6.00$, df $= 3$
6. Binomial Test $N = 8$, $\chi = 6$, $P = \frac{1}{3}$, $Q = \frac{2}{3}$; reject H_0 at $\alpha = 0.05$
7. $\chi^2 = 1.66$, df $= 1$
9. $\chi^2 = 1.789$, df $= 1$
11. $\chi^2 = 6.00$, df $= 3$
12. $\chi^2 = 3.60$, df $= 3$
13. $z = 1.327$, $\chi^2 = 1.761$

Chapter 18

1. a) Sign test $N = 10$, $x = 3$ b) $T = 10$
   c) $U = 23$, $U' = 77$
2. $U = 9.5$, $U' = 62.5$
3. $U = 18$, $U' = 103$
4. $T = 11.5$
5. $T = 48$

INDEX

ABCDE7987654321